May 2018

Breaking the Language Barrier: Equitable Assessment in General Chemistry

Eshani Lee
egandhi12@gmail.com

BREAKING THE LANGUAGE BARRIER:

EQUITABLE ASSESSMENT IN GENERAL CHEMISTRY

By

Eshani N. Gandhi-Lee

Bachelor of Arts - Social Sciences
University of California, Irvine
2005

Master of Arts & Science
University of Nevada, Las Vegas
2012

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy – Chemistry

Department of Chemistry and Biochemistry
College of Sciences
The Graduate College

University of Nevada, Las Vegas
May 2018

UNLV | GRADUATE COLLEGE

Dissertation Approval

The Graduate College
The University of Nevada, Las Vegas

May 7, 2018

This dissertation prepared by

Eshani N. Gandhi-Lee

entitled

Breaking the Language Barrier: Equitable Assessment in General Chemistry

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy – Chemistry
Department of Chemistry and Biochemistry

MaryKay Orgill, Ph.D.
Examination Committee Chair

Kathleen Robins, Ph.D.
Examination Committee Member

Clemens Heske, Ph.D.
Examination Committee Member

Christine Clark, Ph.D.
Graduate College Faculty Representative

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

ABSTRACT

Because language provides the framework through which knowledge is constructed, it is crucial to consider the ways in which students with limited English proficiencies are able to express their understanding. English language learners (ELLs) make up a significant portion of the student body in the education system and represent many ethnic and racial minorities in STEM (Science, Technology, Engineering & Mathematics) fields (Burke & Mattis, 2007). Despite the national push to build a more diversified, STEM-ready workforce, there is little research that considers the way ELLs are assessed in STEM courses at the postsecondary level. Literature reports that science tests that assess the knowledge of students who are still in the process of learning language skills are inadequate and threaten the validity of assessments. The way students interpret and respond to test items are mediated by linguistic and cultural factors, such as home language and prior educational experiences in the country of origin. Therefore, language and cultural factors must be taken into consideration in order to improve the validity of classroom assessments in science courses.

Students' experiences in introductory science courses, such as biology and general chemistry, are critical in their choice of staying in or switching out of STEM majors (Astin & Astin, 1992). Of these, general chemistry is one of the most feared science courses for undergraduate students (Carter & Brickhouse, 1989), and it is a required course for many STEM-bound career paths. Most students struggle with understanding chemistry and many do not succeed on chemistry assessments (Woldeamanuel et al., 2014). Research suggests that scientific language literacy has a significant influence on all students' success in chemistry assessments, including that of both ELLs and Native English Speakers (NES) (Woldeamanuel et al., 2014). Therefore, one way to support the success of all students—and particularly of ELLs—on

chemistry assessments is to address the linguistic complexity inherent in chemistry assessment questions.

One way to ease the burden of linguistic complexity during testing is to apply the Equity Framework of Classroom Assessments (EFCA) (Siegel, 2008) to written test items. This framework aims to make test items more accessible without simplifying the content. In general chemistry, the EFCA can be implemented to make commonly-used items more accessible to all students using modifications such as division of prompt into smaller parts, reduction of non-essential information, adding representation, and simplifying sentence structure.

This study investigated the perceptions of ELL and NES students about general chemistry assessment items that were modified according to the EFCA. ELL students reported to experience difficulties understanding items that included complex linguistic features such as complex sentence structures and vocabulary. The results show that ELLs perceived language-independent features of items to be the most helpful on assessment items. These features included the formatting of items and the visual representations embedded in items. Although NES students also found the visual features of items to be helpful, they used language-dependent features to understand and set up the problems.

The results suggest that ELL students particularly benefited from scaffolding-related features in assessment items. Features that provided content support and guidance for identifying key information and setting up the problems were more helpful for ELL than NES students.

Both groups of students found features that provided contextualization in the form of storylines and/or background information which were not directly related to solving the assessment items to be irrelevant, challenging, and/or confusing.

Both groups of students reported that they preferred the revised versions—which included the modifications recommended by the EFCA—over the original versions of the assessment items presented to them. The findings suggest that most of the modifications employed in the EFCA are effective in mitigating linguistically complex elements of written assessments items about limiting reactant and percent yield in general chemistry and support the assessment of both ELL and NES students.

ACKNOWLEDGEMENTS

This work has been a transformative and arduous process, which would not have been possible without the support of my family, friends and mentors. I would first like to thank my husband and my rock, Eric Lee, for his constant love, support, and devotion. Although there have been many major milestones in this journey, I found the real triumphs to be in the small moments he and I share together, which help me slow down and reflect on the positive aspects in all things. Eric leads by example on how to be resilient, persistent and stand for others who cannot stand for themselves. His encouragement and unfailing support have been essential in this process. Thank you for not allowing me to give up on myself and for always seeing my true potential, especially when I could not.

I would also like to thank my father, mother and brother for believing in me and for instilling in me a sense of ambition and a strong work ethic. As immigrants, we were tasked with the job of survival in an unfamiliar world and significant sacrifices were made for the hope of a better future. Rohan, being your big sister enabled me to set high standards for myself so that I can encourage you to do the same. My father has always believed that I am capable of achieving any goal in life. It is his unwavering faith in me that gave me an unshakeable foundation to persist, accomplish goals, and overcome adversities. My mother has been a tower of strength in this process. She is the voice in my head that tells me that "I can and I will." She is a continuous reminder that this work symbolizes victory for the many women who have come before me and who have not received the encouragement, opportunity and freedom to pursue their dreams. What a luxury it is to search for purpose, meaning, and fulfillment. Mom and Dad, thank you for fighting for me so that I can have this privilege.

I would like to also thank the truly heroic women in my life who understand the cost and responsibility of this privilege. Schetema Nealy has been with me on this journey from the beginning and our bond serves as a true testament to the power of sisterhood in this process. We may have started as colleagues, but over the years, she has been my constant friend and mentor, sister and cheerleader. I cannot imagine going through this journey without someone as brilliant, loving, kind, and loyal. Schetema, you have and will continue to play an integral role in all my accomplishments. Thank you for pushing me to the finish line and beyond. I would also like to thank my dear friend, Jenny Lazarus, who goes out of her way to regularly remind me to trust and believe that all things are possible with faith. Jenny, you are inspirational and a source of my courage. I would like to thank my mother-in-law and friend, Darlene Lee, who has always shown tremendous support of my work. Her inquisitiveness and thirst for knowledge is infectious and serves as an enduring reminder that it all starts with curiosity about the world and the determination to change it.

I would like to extend my gratitude to a great group of colleagues in the Chemistry Education Research group. Thank you, Nicole Baldwin, for your optimism and insights, which have truly added to this project. Thank you, Sarah Wood and Sabrina Barakat, for listening to my talks, providing your important feedback, and much needed laughs.

I would also like to thank the members of my Advisory Committee for your support of and input in this project. Dr. Christine Clark continues to go above and beyond the call of duty to ensure my success. Thank you for imparting your valuable wisdom, knowledge, and enthusiasm for my work. To Dr. Kathy Robins and Dr. Clemens Heske, you have given me encouragement, needed critique, and challenged me to think beyond the scope of chemistry education research.

Lastly, I would to like to express my immense appreciation for my advisor and mentor, Dr. MaryKay Orgill, for her wisdom, guidance, and diligence. Over the years, your insights and encouragement helped me develop my skills and knowledge as a researcher, an instructor, and a student. You have shown me how to think beyond the conventions, take risks, and be innovative. You motivated me to set and meet ambitious goals for myself. Thank you for giving me the freedom to follow my passion on this journey. Your mentorship and commitment provided the solid foundation I needed to learn, grow and succeed.

This work is dedicated to my daughter,

Elara J. Lee.

You are my joy,

my reason to be better today than yesterday,

my sun and my moon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

Project Rationale

Test scores are more than just numbers. They can determine the trajectory of students' lives. For immigrant students, test scores could mean living up to family aspirations in the new country of residence. For those who dream of being the first in their family to get into college, they could mean changing the legacy for future generations. For those who dream of becoming scientists, test scores mean early admissions, waiting another year, or reconsidering the hope of graduate school altogether. For me, test scores have meant all these things at various stages of life.

My family and I moved from India to the USA in 1992. Like most immigrant families, my parents were in pursuit of a better life; but we did not fully anticipate how difficult the transition would be. I had no formal educational experience in English, which meant that I was placed in ESL (English as Second Langauge) courses through the beginning of high school. Navigating the cultural and language barriers in school was a constant battle, especially in classes that did not recognize the issue of English language fluency. I found myself struggling to keep up with verbal instruction and became reluctant to participate in class. The frustration came from knowing that I had the ability to learn the material, but lacked the customary method of communicating my ideas. As a first-generation immigrant student, failing in school was not an option, as my family depended on me to be their delegate and set a positive example for our community.

As an English language learner, my relationship with assessments has generally been exasperating. I remember feeling a fascination with the microscopic world at an early age. I have been a curious student of science since then and have held a deep appreciation of all science-related matters. I recall grasping the concept of photosynthesis clearly in grade school, being able to explain it to my peers, and feeling optimistic about the upcoming exam. Unfortunately, I struggled with deciphering what test questions about photosynthesis were asking of me in my school classes, and my test scores often reflected it.

My struggle during testing became such a frequent pattern in my chemistry classes that it made me feel insecure about my ability to understand or succeed in science. During my introductory chemistry classes at the university, many of my peers (who were both ELLs and Native English Speakers [NES]) had either decided to consider non-science fields, or to study superficially for course exams by rote memorization, learning test-taking techniques (and foregoing the pursuit of a deep understanding of the concepts) and/or relying on computational problems. I call the latter strategy "hiding behind the numbers" because as long as we, as students, knew the mathematical operations of a given problem, we could rely on those operations without having to interpret the science content embedded in the questions.

As I progressed through my undergraduate program, I found that these superficial tactics for approaching learning no longer allowed me to succeed. For those of us who persisted to upper division coursework and were ready to enter graduate school, we found that courses were focused on our abilities to critically think and demonstrate our understanding in various ways, including writing reports, reading research articles, and presenting in class, etc. These tasks required using the academic language of science to reason and problem solve. Although the learning in these courses was difficult, my scientific reasoning and language skills developed, which positively impacted my understanding and appreciation of biochemistry.

It was not until I began teaching undergraduate students in biology and chemistry in graduate school that I realized many younger students were experiencing a similar struggle with language skills in their science courses. I was overwhelmed by how many students struggled with understanding the general discourse used in the textbooks and exams, difficulties that the ELL students generally refrained from sharing with their instructors. Most incoming ELL students in college are conditioned to not ask questions and to keep their challenging in language proficiencies hidden. Because I empathized with the challenges that I observed in my students, I shared the challenges I had faced with reading convoluted laboratory protocols as an ELL. I assured my students that I was a safe person to come to for asking any questions that they did not feel comfortable asking in front of the class. I believe that this allowed them to be more open with me about their challenges, which I appreciated. For example, on one occasion, a young, Hispanic female student asked me "What does this question mean by *tarnish*?" during an exam. I asked her to tell me what she thought it meant. She replied, "like… maybe it got trashed somehow." Unfortunately, this definition was not helping her answer the question because this was not the intended meaning used in the chemistry test question, which was asking about tarnishing as a chemical reaction of Ag (silver) reacting with S (sulfur) containing substances in the air. My response was: "Can you think of it as a process – like something happening to Ag if you leave it outside in the air?" She immediately smiled and nodded. I should mention that this incident occurred when I was proctoring a chemistry exam for an instructor in a lecture course for 200+ students, and questions during exams were explicitly discouraged. Unfortunately, this type of environment is typical in many introductory undergraduate science courses and is not conducive to equitably facilitating students from diverse backgrounds (Bajak, 2014).

After countless interactions with students involving language confusion—including not understanding what is being asked of them on questions and/or problems following written instructions in lab—it became obvious to me that ELLs' challenges with language in their chemistry courses should be examined systematically. Using my experiences with students in general chemistry as a guide, I designed a preliminary study that surveyed undergraduate ELL students in general chemistry about their experiences in their course. More specifically, I asked them to report on how their language fluency affected their learning in lecture, in the laboratory, and on course exams. I surveyed 27 ELL students using a quantitative survey and a qualitative questionnaire. My preliminary study confirmed that most ELL students struggle with interpreting exam questions in general chemistry. Accordingly, I decided to focus my research on examining general chemistry students' understandings of assessment questions. Instead of focusing exclusively on ELLs' interpretations, I chose to focus on how both ELLs and NESs understand chemistry assessment questions. If tests are intended to be reliable tools for assessing knowledge, then *all* students should be given an equal opportunity to demonstrate their understanding.

Research Questions

The impetus for my research came from my personal background as well as the experiences I have had with undergraduate students while teaching general chemistry. I firmly believe that assessments should be conducted in a way that gives all students a fair chance to demonstrate their learning because tests can shape lives. An important goal of this research was to understand how ELL and NES students comprehend general chemistry questions. I was specifically interested in having the students identify features of the questions that they find either helpful or difficult in aiding their interpretations of the questions. The Equity Framework for Classroom Assessment (EFCA) designates specific types of modifications that can make assessment items more accessible for all students. The current study examines whether the modifications suggested by the EFCA are useful for decreasing the linguistic complexity of university-level chemistry assessment questions. The following research questions will be investigated in this study:

1. What are English language learners' (ELLs) and native English speakers' (NES) perceptions of typical general chemistry exam questions as compared with chemistry exam questions that have been modified according to the equity framework for classroom assessments (EFCA)?
    a. Which features of the questions do ELL and NES students perceive to be helpful?
    b. Which features of the questions do ELL and NES students perceive to be challenging?
2. What (additional) modifications do ELL and NES students believe would make chemistry exam questions easier to comprehend?

CHAPTER 2 LITERATURE REVIEW

Overview

In this chapter, I review the existing research literature on assessing ELLs in chemistry. Figure 1 depicts the most relevant topics of discussion represented in the overlapping areas of the diagram. Because there is a limited amount of literature focused exclusively on assessing ELLs in chemistry, this literature review will focus on more foundational areas, including literature related to (1) ELLs in general, (2) the assessments of ELLs, and (3) assessments in chemistry. The intersection of these three bodies of literature represents the main area of focus of this study, which is indicated by the star symbol in the center of the diagram in Figure 1: studies that focus on the assessment of ELLs in general chemistry. Each of these sections is further divided in subcategories that are represented in corresponding flow charts. Given the limited amount of literature focused on the assessment of ELLs in university-level general chemistry courses, these sections have been expanded to cover ELLs in science and the assessment of ELLs in science. This review underscores the gap in literature about the assessment practices of ELLs in chemistry. My study addresses this gap.

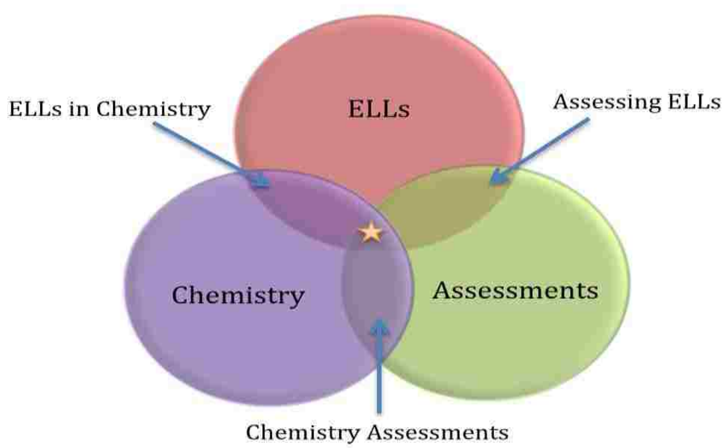


Figure 1. An overview of the bodies of literature included in this study.

English Language Learners

In this section, I begin with a brief discussion of how ELL populations have been defined historically. Then I review literature regarding how ELLs are defined in postsecondary education, which precedes a discussion of prevalent language acquisition models in the field. This section also includes research on the academic language in science and how it pertains to undergraduate ELL students. Figure 2 represents the organization of this section.

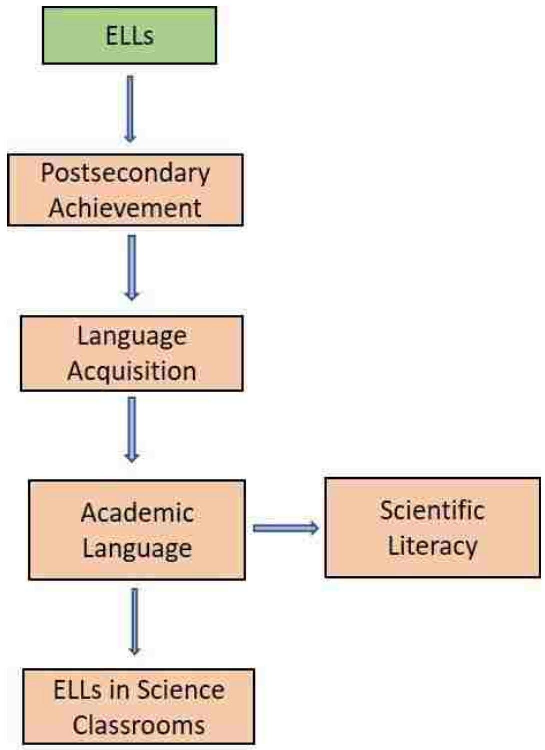


Figure 2. An overview of subcategories under the "ELLs" section.

The English Language Learner Population

Historically, it has been challenging to identify this population of students because of the population's inherent heterogeneity and linguistic diversity. ELLs represent various different language backgrounds as well as different socioeconomic statuses, academic experiences and immigration histories. Consequently, there are many terms used to identify ELLs, including the following common terms: ESL (English as Second Language), CLD (culturally and linguistically diverse), LEP (limited English proficient), EFL (English as a foreign language), NELB (non-English language background) (American Institutes for Research, 2012). For the purposes for my study, I use the term English Language Learners (ELLs) as it encompasses the target population for my study.

No matter the definition employed, the U.S. educational system has seen considerable growth in the ELL population over the past decade. In the 2012-13 year, 9.2% or 4.4 million students were identified as English language learners in U.S. public schools. An ELL student is defined as a student "whose primary language is not English and whose English proficiency is below the average proficiency of peers whose primary language is English" (Education Commission of the States [ECS], 2014). Using this definition, Nevada has one of the largest ELL student populations in the country (National Center for Education Statistics [NCES], 2012). Additionally, the ELL student population in Nevada increased at a rate of 35% between 2002 and 2012, more than three times the rate growth of the ELL student population in the country as a whole (Mokhtar, 2012). At the local level, Clark County has experienced a 19.6% increase over the same time. As of 2012, there were approximately 84,125 ELL students enrolled in the Clark County School District (ECS, 2014).

Postsecondary Achievement of ELLs

The challenging process of identifying ELL student populations has made it problematic to gain information about students' achievement across school districts and institutions. Generally, there is a shortage of information available about academic performance and classroom instruction because the criteria for ELLs' academic progress are different across schools and because certain types of ELLs are exempt from school assessments (Lee & Fradd, 1998). Studies that track ELL students' achievement across grade levels and into college are also rare but necessary in order to understand issues of postsecondary access and achievement.

Kanno and Cromley's (2013) study followed a representative sample of 10,300 eighth graders from 1,052 randomly selected schools across the country for 12 years, with the goal of learning about the cohort's postsecondary attainment and access. The cohort was divided into three categories: English-monolingual students (EMs), English-proficient linguistic-minority students (EPs), and ELLs. EMs were students who were native English speakers, EPs were previous ELL students who were now considered fluent in the English language, and ELLs were students whose English language proficiency was still developing. The dependent variables used were access to and attainment in postsecondary education. Access was operationally defined as students' first college-level institution (bachelor's level, vocational program/community college and/or none). Attainment was operationally defined as students' highest degree earned. Their findings showed that only one in 8 ELLs earned a bachelor's degree, compared to one in four EPs and one in 3 EMs. Additionally, one in five ELLs dropped out of high school.

The results of Kanno and Cromley's (2013) study indicate that there is an achievement gap between ELLs and English proficient students that persists throughout postsecondary education. The results of this study also corroborate government data that show that, compared to English proficient students, ELLs have lower levels of academic achievement and higher rates of poverty, mobility, and high school non-completion (NCES, 2004). On a broader scale, these findings imply that many ELLs are at risk of unemployment because most jobs now require at least some level of postsecondary education (Kanno & Cromley, 2013). Findings such as these encourage a deeper examination of how language acquisition impacts the ability to learn. The topics of language acquisition and its role in learning for ELLs are discussed in the next subsections.

Language Acquisition

The process of language acquisition sheds light on ELLs' capacities to perceive written and verbal instruction. In this section, I discuss two concepts related to language acquisition: the types of language proficiencies and the continuum of language proficiency. A major contribution in the field came from Jim Cummins (1981), whose work suggests that students whose first language (L1) is not English and who have had two to three years of schooling in their L1 need an additional five to eight years in a U.S. school to match their peers' level of English proficiency.

Cummins' later work (2000) explained that there are two types of language proficiencies. BICS (basic interpersonal communicative skills), typically acquired through everyday social interactions, become conversational and informal communication skills. CALP (cognitive academic language proficiency) is more difficult to develop as it takes place in formal, academic settings and involves a higher cognitive load than BICS. Collier (2008) defines academic language as terms not used in everyday conversation, words that tend to have one meaning in everyday talk but mean something else in the classroom, technical terms and higher level concepts.

Cummins (1980) emphasized that BICS refers only to some fundamental, rapidly developed features of communicative proficiency, which are relatively superficial aspects (e.g., accent). Unfortunately, educators may build their judgements of students' proficiency based on BICS alone. For example, a teacher may assume that an ELL student should be able to comprehend written instructions in the same way as NES students because she/he seems to be able to converse with peers in social interactions in the English language. However, CALP could only be developed within a formal academic learning setting. Developing CALP includes listening, speaking, reading, and writing about a given content area. Academic language is much more involved than understanding vocabulary of the content area; it requires higher order skills. For example, comparing, classifying, synthesizing, evaluating and inferring are skills that require a more developed CALP. For immigrant students without prior schooling or support in L1 development, CALP development takes seven to ten years (Collier, 1995).

To conceptualize language proficiency in a way that the developmental interrelationships between academic performance and language proficiency can be illustrated, Cummins (1984) proposed a model (Figure 3) that outlines language proficiency along two continua. The horizontal continuum is based on the range of contextual support available for constructing meaning, described as "context-embedded" versus "context-reduced" communication. In context-embedded learning, students are able to negotiate meaning and language using the wide range of situational cues and feedback available. For example, engaging in discussions, participating in group work, writing a personal letter, or reading an article would be context-embedded communications and would be located on the left side on the continuum. In context-reduced learning, students must primarily rely on the meaning of the words themselves and may need to disregard their "real world" knowledge in order to interpret the logic of the communication adequately. For example, textbook reading, classroom lecture, and specialized vocabulary would be considered as context-reduced communications and would be located on the right side of the continuum (Cummins, 1984).

[redacted]

Figure 3. Cummins' (1984) continuum of language proficiency (Cummins, 1984, p. 31).

The vertical continuum describes tasks that require little cognitive load (top) and tasks that require a higher cognitive load (bottom). Cognitively undemanding communicative tasks occur when linguistic tools have become essentially mastered. At this level, little active cognitive involvement is needed to perform appropriately. For example, a conversation at a coffee shop or simple yes/no question in the classroom requires a low level of mental effort. On the other hand, cognitively demanding communicative tasks require active cognitive involvement to perform appropriately because language proficiency has not become automatized. For example, analyzing and synthesizing information quickly in a chemistry lesson or a multiple-choice test, or writing this literature review would require a higher level of language proficiency (Cummins, 1984). Cummins' works suggest that to aide ELLs in the development of CALP, higher cognitive load tasks can be supplemented with contextual cues in the content. For example, adding illustrations to a multi-step chemistry problem could ease the burden of decoding technical language and translating English.

The Importance & Challenge of Academic Language. Cummins (1980, 1981, 1984) emphasized that the extended timeframe for acquiring proficiency in academic language should not be confused with a *disability to learn*. Educators may attribute poor test scores of ELLs who may sound fluent because of well-developed BICS to laziness or lack of aptitude, failing to recognize the cognitive load of learning in commonly context-reduced areas such as science or math, where there is a higher frequency of technical vocabulary and few descriptive cues. "The most important thing [teachers] can do is to understand what academic language is and how they can teach [ELLs] this type of discourse" (Collier, 2008, p.10).

Generally, science teachers are not well-prepared to help their students understand the complexity of academic language. Bruna et al.'s (2007) study examined how a 9th grade teacher's own conceptualization of academic language influenced her teaching in a science classroom, which included many ELL students. Their observations suggested that the teacher's instruction on the rock cycle mainly focused on discrete vocabulary, not other linguistic resources such as grammar, lexical items and semantic relations in the process of rock formation. This type of instruction "takes, as we see it, the 'motion' out of meaning making" (Bruna et al., 2007, p. 46). The findings highlight that a simple focus on discrete vocabulary is not sufficient when trying to help students understand academic language (Bruna et al., 2007).

Unfortunately, teaching only the meaning of key words without explaining the use of other general-purpose academic words that students do not know could impede student understanding (Snow, 2010). This is especially the case for ELLs who do not recognize that these terms are being used in technical ways with different meanings from those of everyday life (Snow, 2010). For example, the definition of torque is "the product of the magnitude of the force and the lever arm of the force" (Snow, 2010, p. 452), but many students are also unfamiliar with the terms within the definition such as "force," "arm" and even "product," which often puts them at a disadvantage when it comes to following classroom instruction clearly.

Scientific literacy. Research suggests that academic language makes acquiring scientific literacy challenging, especially for those who are still in the process of learning the basic rules of the English language (Bruna et al., 2007; Collier, 2008; Cummins, 1980, 1981, 1984; Lemke, 1990; Snow, 2010). Those who are scientifically literate have the knowledge of scientific concepts and processes needed to make informed decisions, participate in society, and contribute to economic productivity. According to the American Association for the Advancement of Science (AAAS), science learning involves a two-part process "to acquire both scientific knowledge of the world and the scientific habits of the mind at the same time" (AAAS, 1989, p. 190). Scientific knowledge development consists of knowing, doing, and talking science (Lee & Fradd, 1998). In this section, I will discuss the prevalent work of Lee and Fradd (1998), which underscores the significance and challenges of developing scientific literacy for students of non-English-language backgrounds.

For ELLs, knowing, doing, and talking science each impose unique difficulties that are often not recognized in the classroom and that hinder ELLs from achieving academic literacy in science. Lee and Fradd (1998) review each of these aspects of learning science from the perspective of ELLs. *Knowing* science is developing scientific understanding, which requires building new knowledge on the foundation of prior knowledge, using science vocabulary, and understanding concepts and relationships. Because the knowledge that each student brings to the classroom differs, teachers should help students connect their prior knowledge with newer concepts they are expected to know. This can be particularly useful for ELL students, who may have learned a concept differently than the way it is taught in the U.S. For example, most students from non-English-language background could be more familiar with the metric system of measurements than with the U.S. system of measurement. In a science classroom, the ELLs' prior knowledge about the metric system could be leveraged to help the students develop an understanding of the U.S. system of measurement.

Learning vocabulary is another important aspect of knowing science that is not as straight-forward as memorizing a list of terms. Words of one language cannot always be directly translated to another language; hence, meanings of words must be understood within contexts. For example, the word *logic* is not directly translatable in Chinese and Hindi (Lee & Fradd, 1998). ELLs struggle with using the vocabulary words in the same frequency and manner used in the context of science; this is often manifested as students saying too much or too little, giving the impression that they do not understand the content when the issue is most likely that they lack the language or communication patterns to express precise meanings expected in the science classroom. For example, a student with limited verbal proficiency in English might use extra filler words such as "stuff" or "things" to substitute for accurate scientific words of which they may not be aware (Lee & Fradd, 1998).

Doing science is thought of as scientific inquiry, or engaging in inquiry and solving real-world problems (Lee & Fradd, 1998). The Next Generation Science Standards emphasize asking questions and defining problems as one of the most important practices of science (Next Generation Science Standards, 2014). Students engage in inquiry by making observations, proposing explanations, interpreting and verifying evidence, synthesizing ideas to make sense of the natural world, and manipulating materials. Many aspects of inquiry are difficult for all students because they require language functions, such as reflecting, predicting, inferencing, and hypothesizing. For ELLs, there is an added challenge of having different oral traditions at home and/or cultures of prior education systems where asking questions or devising plans for investigation on their own may not have been encouraged. This is particularly the case for students from cultures that are taught to respect authority; they have been taught to listen to teachers and to let teachers direct them, rather than to inquire, explore, and seek alternative ways (Lee & Fradd, 1998).

Talking science is central in scientific discourse, which provides the structure for participating in social and academic discourse, using multiple representational formats, and appropriating the discourse of science (Lee & Fradd, 1998). Talking science means to communicate in the language of science and act as a member of the scientific community (Lemke, 1990). For ELLs, communication patterns at school are different from those at home. Because of this, they have different interpretations of verbal communication and expression. For example, many ELLs rely on the use of gestures to supplement and replace words more than monolingual English-speaking students (Michaels & O'Conner, 1990). When asked to describe the concept of "balance" in class, a Haitian female student, who understood the concept, was unable to express it using the discourse pattern of "why-because" and could not make her mental operations explicit in her verbal explanation. Instead, she used hand gestures to aid her description. Consequently, her responses were sometimes seen as less intelligible than those of other students (Michaels & O'Conner, 1990).

Overall, ELL students face unique challenges as they attempt to develop scientific literacy. These challenges potentially put ELLs at a learning disadvantage compared with their NES counterparts. However, according to research, there are steps that teachers can take to help their ELL students overcome some of these challenges. For example, teachers can make academic content more accessible by providing contextual cues during instruction and activities.

This includes applying content to students' everyday lives, as well as providing applicable examples of concepts where possible. Teachers can also provide hands-on, interactive learning activities and/or use illustrations in their teaching (Lee & Fradd, 1998). Although these steps do not address all of the challenges faced by ELL students in science classrooms, they can at least reduce the gap between the performances of ELL and NES students.

English Language Learners in Science Classrooms

In order to further explore the issues associated with knowing, doing, and talking science from the perspectives of ELLs, the following section focuses on the limited number of studies in the literature that discuss ELLs and their perceptions of the science classroom. As a reminder, because of the lack of research about ELLs in chemistry, this category has been expanded to discuss ELLs in science classrooms. Literature suggests that ELL students tend to carry lower confidence in their abilities to succeed in class compared to NES students and often feel isolated in the classroom (LeClair et al., 2009; Ryu, 2015). Additionally, newcomer ELLs, those who arrived in the U.S. less than a year ago, are less likely to feel confident about verbally participating in class, which not only negatively impacts their participation in class, but also their abilities to branch out and make connections with NES students in the class (Ryu, 2015).

LeClair et al.'s (2009) study examined how ELLs view themselves and their peers in the classroom. The study included 257 elementary school (3^{rd} through 5^{th} grades) students' perceptions about classroom relationships (teacher-student, peer, and home-school) and self-regulation (self-efficacy, self-determination, and self-control). It compared perceptions of ELL and NES students using a quantitative class maps survey with subscales for classroom relationships and self-regulation. The results showed that ELL students rated themselves as having lower academic self-efficacy and rated their NES peers as having higher levels of self-control in the classroom. ELL students also rated their class as more orderly and their NES classmates as more regularly following rules (LeClair et al., 2009). The findings suggest that ELLs may be aware of their own academic performances in the classroom compared to their NES classmates and that they may not be confident about their ability to perform equally well. ELLs may also be less familiar with the classroom culture and, therefore, think of themselves as less likely to follow the rules of the class.

Similarly, Ryu's (2015) study suggests that unfamiliarity of the school culture influences how ELLs position themselves in the classroom. This ethnographic study examined how newcomer Korean students positioned themselves, participated and learned in two Advanced Placement (AP) biology classes. The research questions focused on how the students evaluated, perceived, and positioned each other in the class in terms of Biology achievement and verbal participation. "Good" students in the class were considered to be those who adopted pieces of biological knowledge presented by the teacher and performed well on class exams. Top students were called out for receiving top scores. Accordingly, students who did not perform well were positioned at a lower status in the social structure of the class. Class achievement shaped group positioning in the class such that higher achievers often formed groups with students who achieved similar or higher scores than them. This resulted in most newcomer Koreans forming separate groups from other students of racially and linguistically dominant groups (Ryu, 2015).

Verbal participation also seemed to be an important factor that influenced students' positions in the classes. The teacher believed that adopting "biology language" was an important part of learning biology and encouraged her Korean students to translate biology content into Korean and help each other "talk biology." However, many of her newcomer Korean students remained reluctant to verbally participate. Students stated the following when asked about their participation, “I could do [participate] better if I spoke English better” or “It’s embarrassing [to ask questions in class] because I ask when I don’t understand English (p. 359).” Because of newcomer Korean students’ infrequent verbal participation in whole class settings and their speech patterns being perceived as foreign, they were often positioned at a lower level than native English speaking students by peers in their classes (Ryu, 2015).

This study highlighted the fact that the Korean students’ status as newcomers in the U.S., combined with their limited ability to verbally participate in the class contributed to their feeling of benig unsuccessful and disempowered in their biology class. These findings brought to light how the expression of knowing and talking science can often be narrowly defined in the science classroom. Many students perceived a pressure to provide one correct answer on exams and to the teachers, instead of demonstrating their unique understanding of scientific phenomenon in alternate ways. Students also perceived that they would be thought of as unintelligent or disregarded if they were not able to ask and answer questions in class using the type of discursive patterns (or academic language) accepted in science (e.g., factual information with a heavy use of technical vocabulary) (Ryu, 2015).

Summary. Reports in literature contend that the ELL student population is on the rise in the public school system. However, ELL students are less likely to earn bachelor’s degrees and are more likely to drop out of high school than their NES counterparts. Research in the area of language acquisition suggests that ELL students who sound fluent may still be struggling with grasping the academic language of a discipline, especially in the case of science. In turn, according to the literature, underdeveloped academic language proficiency can be a barrier to scientific literacy. It should be noted that most research on ELLs has been exclusively focused on the language acquisition of school age ELL children. It is not yet clear whether the results of this research apply to adult ELLs.

Assessing English Language Learners

Because an important goal of the current study was to examine how students, including ELLs, respond to general chemistry assessments, it was critical to review the literature involving the assessment of ELLs. Because of the lack of research specifically about the assessment of ELLs in chemistry, I decided to utilize the following two bodies of research to inform the design of the project: (1) assessing ELLs and (2) chemistry assessments, which are discussed in the next two sections. The discussion of the assessment of ELLs includes research related to the following topics: (1) the obstacles in assessing ELLs' content knowledge, (2) measurement error as related to the assessment of ELLs, and (3) developing equitable assessments. Figure 4 shows the organization of these three sub-categories in this section.

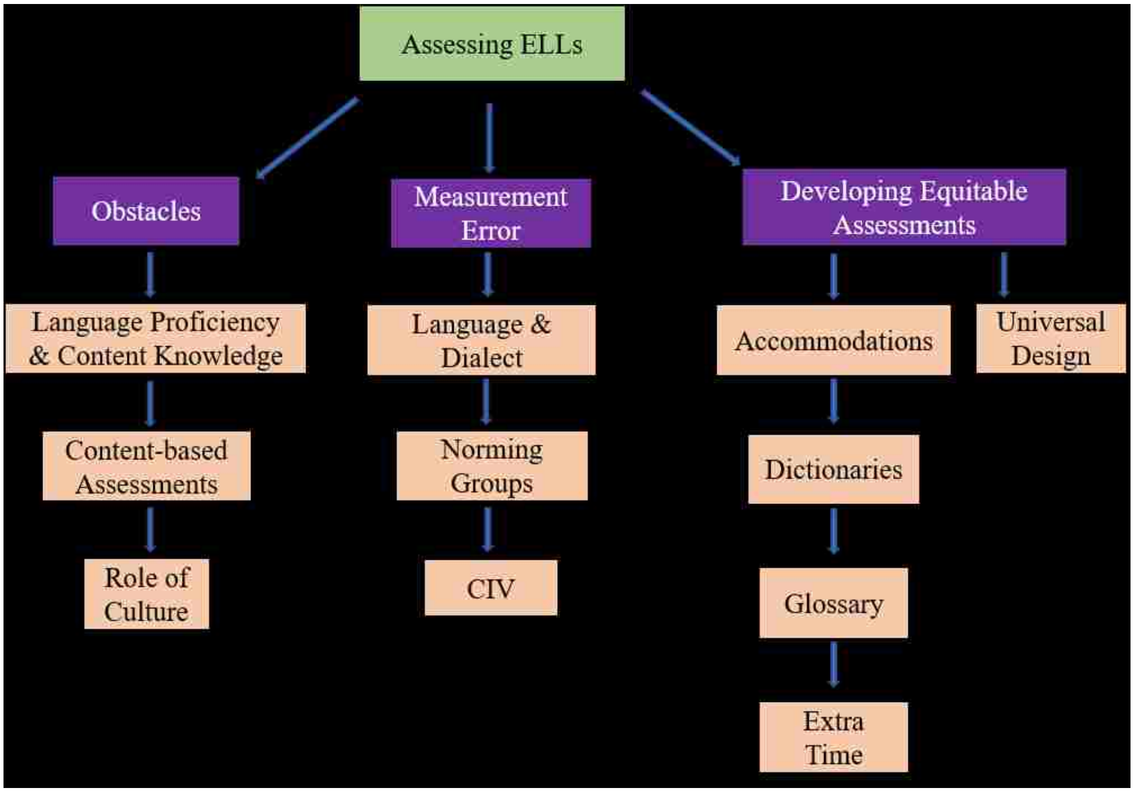


Figure 4. An overview of the "Assessing ELLs" section.

Obstacles in assessing ELLs' content knowledge

The assessment of ELLs is a necessary but challenging endeavor. In this section, I discuss the main barriers in assessing the content knowledge of ELLs. Arguments about effective ways of assessing ELLs' content knowledge center on the relationship between language proficiency and test performance. How well an ELL student can read in English is an important factor in determining how she can be assessed. "Because language plays an integral role in most, if not all, academic learning, any test of academic achievement is also, to some degree, a test of language ability" (Kieffer et al., 2009, p. 1170). Key obstacles, as identified in literature, in the assessment of ELLs fall under three main topics: (1) the relationship between language acquisition and content knowledge, (2) academic language in content-based assessments, and (3) the role of culture in assessment (de Schonewise & Klingner, 2012; Teaching English to Speakers of Other Languages [TESOL], 2008).

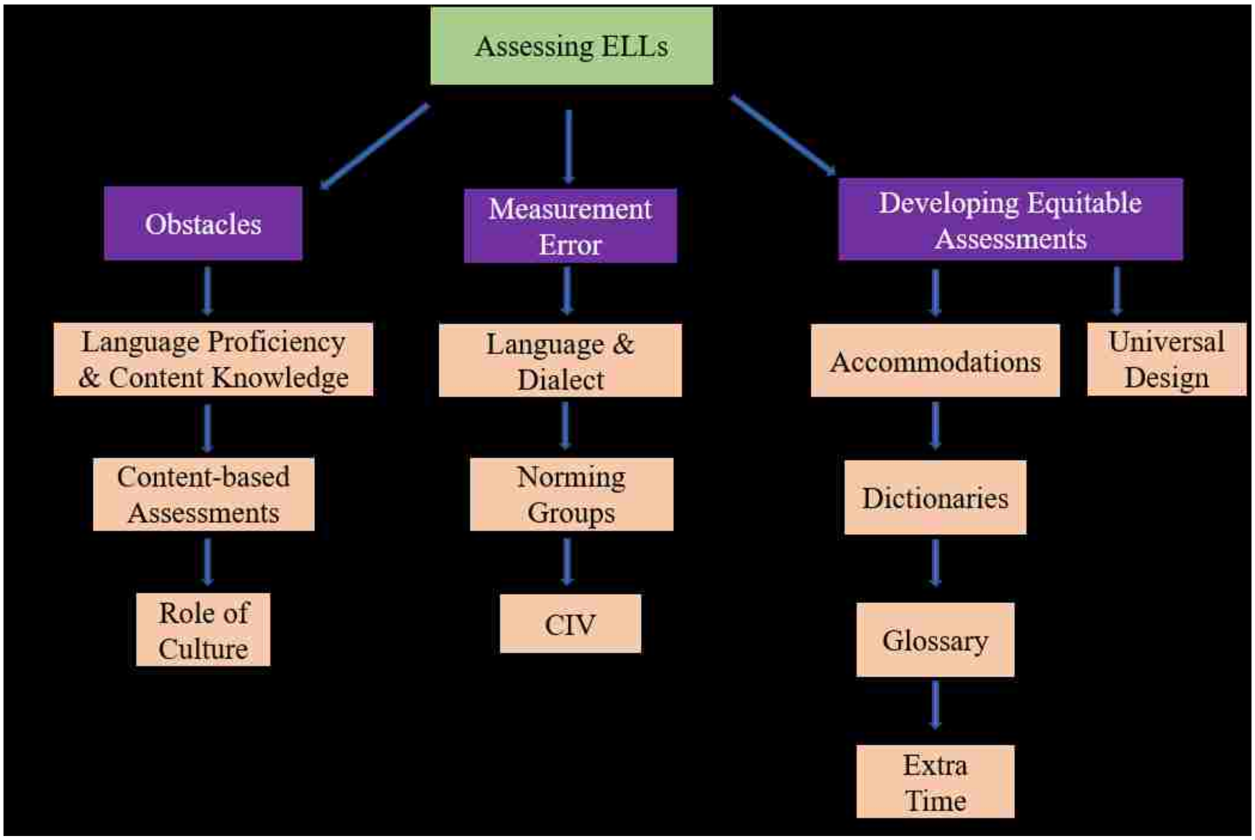


Figure 5. An overview of the subcategories under the "Obstacles" section.

Relationship between language proficiency and content knowledge. Research indicates that because of the interplay between language proficiency and content understanding, ELLs encounter different and additional challenges when they attempt to make sense of math and chemistry test items. These challenges are outlined below:

- Because English is often their second language, ELLs' math knowledge needs to be filtered through English, which means math operates as their "third" language. On exams, ELL students have to try to navigate highly demanding cognitive tasks and interpret abstract and complex ideas across language platforms in a timed setting (Chamot & O'Malley, 1994).

- Similar to math, sciences such as chemistry are also comprehensive in nature, and learning is accrued. For example, students must have knowledge of the periodic table of elements before they can understand how elements bond with each other (Chamot & O'Malley, 1994). Similarly, if ELL students develop incorrect understandings about words or phrases used early during the semester, they may develop incorrect understandings of concepts presented later in the semester. For example, not understanding what "stability" means would impact the way they understand the rationale for drawing energy diagrams based on electron stability.
- Some of the terms used in math and chemistry are words that are not used in everyday talk or that can have different meanings than they have in everyday talk. Additionally, vocabulary tends to be technical, specialized and decontextualized on test items in math and chemistry. For example, words such as *stoichiometry*, *quotient*, and *product* can be unfamiliar and can have different meanings in the context of chemistry. ELLs are more likely to get word problems wrong that contain large amounts of technical vocabulary (Abedi & Lord, 2011).
- The syntax on math and chemistry items does not always resemble basic language structures used in everyday talk and can, thus, be cumbersome and confusing. For example, passive voice (e.g., X was reduced by Y), comparatives (e.g., lower than, greater than), and unknown variables (e.g., X is the product of Y less than 2) can be perplexing to ELL students (Chamot & O'Malley, 1994).

- Without sufficient contextual cues built in, readily interpreting chemical expressions embedded in math can also be extremely perplexing for students, including ELLs (Chamot & O'Malley, 1994).

The focus of the current study is on students' interpretations of chemistry assessment items. While this is a challenging task for all students, it is potentially more complex for ELLs than for NES students. Figure 6 shows a comparison of potential steps taken by NES and ELLs when solving chemistry word problems. (Note: This diagram has been included for the purposes of visual clarity, and is not intended to imply that problem solving processes are linear in nature.)

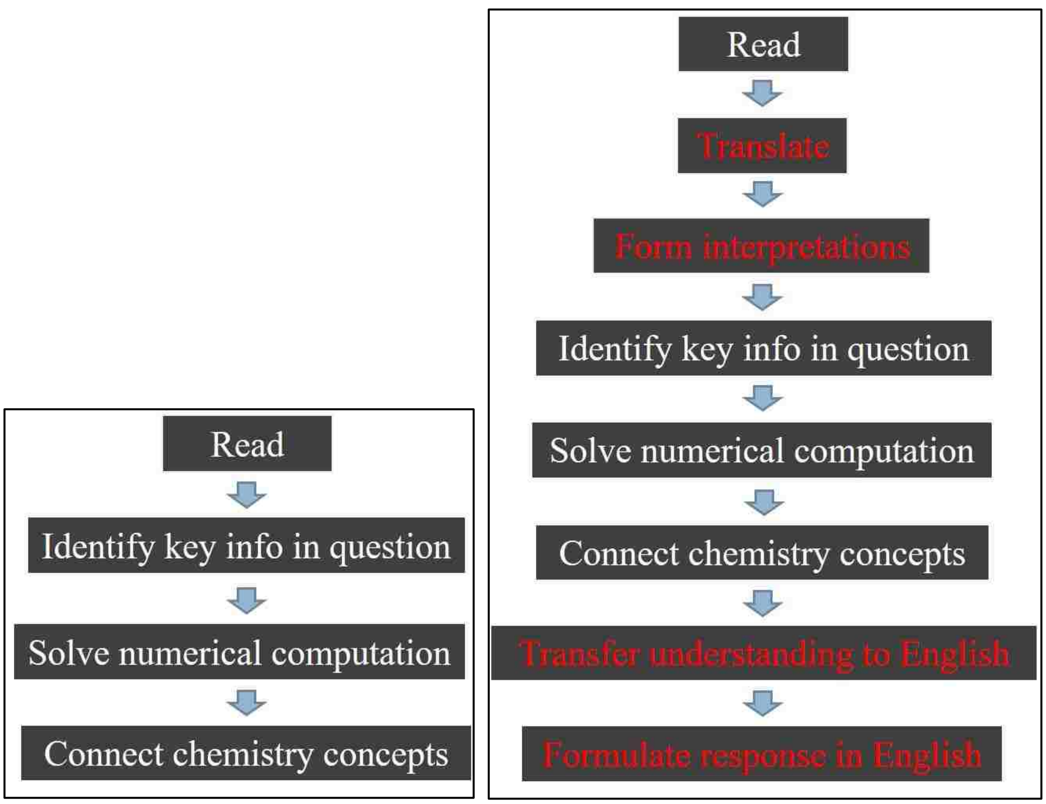


Figure 6. An outline of processes students take to solve chemistry problems involving computations. Left: Steps that all students go through when interpreting chemistry test questions. Right: Additional steps that ELLs must take in order to interpret chemistry test questions (highlighted in red).

All students must read the question first; then identify the most relevant pieces in the question; then perform numerical computations to get a numerical output; and, lastly, express this number in terms of the context of the chemical problem. ELL students must take a few additional steps to solve the same problem: After reading the problem, they must attempt to translate parts of it to their L1, and then form interpretations of what the question may be asking. After connecting the numerical answer to the context of the chemistry problem, they must transfer that answer into the English language and express their final answer in English. According to the literature, word problems embedded in science content, such as in chemistry, are especially taxing for ELLs, and the added steps ELLs must take to solve chemistry problems often result in the ELLs getting these problems wrong on assessment tasks (Abedi & Lord, 2011).

Content-based assessments of ELLs. Several studies have examined the role of language proficiency on standardized, content-based assessments (such as science and math tests) and found that the tests may inadvertently function as English language proficiency exams for ELLs (Abedi, 2002; Brown, 2005). Abedi's (2002) study examined the impact of students' language background on the outcome of achievement tests, using data collected from four different K-12 schools across the U.S. The analyses were focused on comparing the levels of performance of ELL and NES students by multiple-group factor analysis of test items. The results indicated that the overall means for ELLs' test scores were significantly lower than those of NES students in content areas with higher language demands. The performance gap between ELLs and NES students tended to be smaller in lower grades (e.g., Grade 2) and larger in higher grades (e.g., Grade 9). The results suggested that language background adds a compounding factor, referred to as the “language factor,” to the assessment of ELLs, which may be a source of measurement error (Abedi, 2002).

The issue of the “language factor” is of particular concern in evaluating students’ math skills. Although there is a growing emphasis on the need for all students to develop strong math skills in school, the way math skills are evaluated using standardized assessments combine ELLs with all other students to be evaluated in the same manner, which could put ELLs at a disadvantage. Math-based tests that require students to read complex, multiple-part questions and read and/or provide written responses describing their explanation and problem-solving process are considered particularly unfit and problematic for evaluating a student population still acquiring English language skills (Brown, 2005). Brown’s (2005) study examined the Maryland State Department’s (MSDE) literacy-based performance assessment designed for math testing in Grades 3, 5, and 8 to find differences in scores between ELL and NES students within the same socioeconomic status (SES). The analysis found that high-SES NES students outperformed high-SES ELLs, but there was no significant difference found between low-SES NES students and low-SES ELLs. Results suggested that high-SES ELLs may have scored lower because their language background impeded their performance on the math-based test. The researchers speculated that ELLs’ true math ability could be concealed by their under-developed academic language proficiency (Brown, 2005).

Language also becomes a factor in assessing students’ content knowledge in science courses. Noble et al.’s (2012) study claimed that large-scale, standardized science assessment differentially measures students’ knowledge. They examined 36 students’ responses on multiple choice questions (MCQs) on a standardized science test for 5th, 8th and 10th grade students. The participants included low-income students, ELLs, and middle-class NES. Researchers found that low-income and ELL students were more likely than middle-class NES students to answer science assessment items incorrectly despite demonstrating knowledge of the corresponding science content during interviews.

Overall, the studies discussed above suggest that ELLs tend to answer test items differently than NES students, but not because of a lack of knowledge or ability. This evidence refutes the assumption underlying the psychometric model of large-scale tests that if a student incorrectly answers questions on the standardized test it is because the student lacks proficiency in the content knowledge (Noble et al., 2012). Student performance on math or science assessments is not solely a function of the math or science content they were taught or their resulting level of knowledge. It is also dependent on students' culture and language, factors which are not often considered in the interpretation of student performance on science tests (Messick, 1989). The major focus of the current study is the influence of language on students' interpretation of assessment items in chemistry. However, because language cannot be completely divorced from culture, I briefly discuss the role of culture in assessing ELLs' knowledge in the section that follows.

Role of culture in assessing ELLs. Given that linguistic and cultural factors are closely intertwined, the issue of cultural validity has become an important point of discussion in the literature as it pertains to assessing ELL students' content knowledge. Culture shapes the way we construct knowledge and form meanings from experience, which is linked to the way we think, reason, and solve problems (Solano-Flores & Nelson-Barber, 2001).

> Because culture and society shape mental functioning, individuals have predisposed notions of how to respond to questions, solve problems and so forth. It follows that these predispositions influence the ways in which students interpret material presented in tests

> and the ways in which they respond to test items. Surprisingly, this view has not been incorporated into the set of actions required to develop valid assessments. Current approaches in assessments give little consideration to understanding how these sociocultural predispositions influence student thinking. (Solano-Flores & Nelson-Barber, 2001, p. 554)

The resistance to integrating cultural validity into instruction and assessment is especially prevalent in science and mathematics; however, integration is indispensable if the goal is to make science accessible to all (Lemke, 2001; Luykx et al., 2007). Traditionally, science and math are considered bodies of knowledge that are universally valid and "culture-free," in that they remain the same regardless of social and cultural groups that take part in it. This conception of science in incompatible with a multicultural approach to science education. Lemke (2001) argues that science is a social enterprise and, even if we ignore the large-scale society that science operates in, we cannot ignore that students' attitudes toward science, beliefs, and identities are a consequence of a student's life outside the classroom. Classrooms are not closed communities. He suggests that an overly narrow and rationalistic view of science is unwelcoming to the new era of students trying to approach science in a new global economy (Lemke, 2001).

Currently, the assessment of linguistic (and cultural) minority students is guided by superficial assumptions about language, cultural misconceptions, and stereotypes rather than a multicultural approach to science (Solano-Flores & Nelson-Barber, 2001). Research in science education on assessment tends to focus strictly on cognitive psychology to explain performance tasks without taking into consideration the social and cultural aspects of science teaching and learning. This has been attributed to the assimilationist perspective in science, which assumes that others should adapt to the Western way of knowing and doing science (Solano-Flores & Nelson-Barber, 2001).

However, studies that examine how knowledge is acquired and organized assert that the cultural component is salient. For example, research suggests that categorization of objects and ideas is culturally dependent. Bilingual Chinese organize objects more relationally than European Americans: by the internal association of the objects to each other instead of the identities and functions of the objects. In terms of reasoning skills, research indicates that the Chinese reason in a holistic and relation manner compared to European Americans, whose reasoning skills are more analytic (Ji & Zhang, 2004).

Solano, Nelson, and Trumbull's (2003) study illustrates that the same question can be interpreted differently depending on the test taker's cultural background. The following three groups of 4th and 5th grade students were presented with the National Educational Assessment Progress (NAEP) mathematical test problems: (a) White, suburban, high income; (b) American Indian, rural, low income; and (c) African American, inner city, low income. The test question referred to as the "Lunch Money" item, is given below:

> Sam can purchase his lunch at school. Each day he wants to have juice that costs 50 cents, a sandwich that costs 90 cents, and fruit that costs 35 cents. <u>His mother has only $1.00 bills</u>. What is the least number of $1.00 bills that his mother should give him so he will have enough money to buy lunch for 5 days? (Solano, Nelson & Trumbull, 2003, p. 4; underlining added for emphasis)

The researchers found that 84% of White students interpreted the underlined sentence as intended by the test designer. The intended meaning of this phrase is that his mother has $1 bills, but does not have any coins or any other bills. However, only 56% of American Indian and 52%

of African-American students read the sentence as intended, with 10% and 18% of American Indian and African-American students, respectively, interpreting the word *only* as limiting the number of dollars and not the number of dollar bills. The follow-up questions during the interview with a low-income, minority student revealed that he thought about the context of this problem differently than intended:

> Researcher (R): Now, what do you think this question is asking from you? What is it about?
>
> Student (S): It's about Sam and he wants to buy his juice, his sandwich and his fruits. For lunch. Maybe he was hungry. But, I think his mom didn't have enough money.
>
> R: Why?
>
> S: Because she only had one dollar bill. (Solano, Nelson & Trumbull, 2003, p. 5)

The analysis of such items in the study showed that the construction of interpretations of the wording on the test were influenced by cultural backgrounds; however, because culture is not generally factored into test performance, these issues go largely undetected for linguistic minority students on standardized exams (Solano-Nelson & Trumbull, 2003).

Another study that examined how students' prior knowledge and cultural knowledge shaped their responses on science assessments found that tests are not culture-free entities (Luykx et al., 2007). The project studied the responses on 6,000 tests (two pre-tests and two post-tests) administered to 1,500 3rd and 4th grade students on the topics of measurement and matter (3rd grade students) and the water cycle and weather (4th grade students). The study conducted a qualitative analysis of the influences of the students' home language on their written responses to the administered questions. The researchers also looked for evidence of the influences of cultural beliefs and linguacultural factors (e.g., voice, genre, and framing of responses) in students' written responses. The findings are summarized below under the bulleted categories of (1) linguistic influences, (2) cultural influences, and (3) linguacultural influences (Luykx et al., 2007).

- Linguistic influences: Students' responses in English differed in spelling, reflecting the phonology or orthography of the home language. When these responses did not reflect the spelling and/or pattern of Standard English, teachers often regarded the response as incorrect or as indicating a lack of understanding of content knowledge. For example, "…*the waro gos to the nodo baro*" [the water goes to the other bottle] and "Meibi the to spribriment the to or abaut eor" [Maybe the two experiments the two are about air] (Luykx et al., 2007, p. 909). Additionally, students seemed to interpret science terms in terms of their everyday meanings and associations, instead of their specialized meanings in the context of science. For example, the ELLs in the Luykx et al. study (2007) confused *gas* (state of matter) with gasoline, *states* of matter with geographical states, and scientific *instruments* with musical instruments.
- Cultural influences: Home norms, practices and beliefs seemed to surface in many student responses. For example, when asked a question about where condensed water droplets come from, one student answered, "A leak in the roof" while others said "God makes it rain." When asked a follow-up question "where did the [evaporated] water go?" another student wrote "Someone stole it" (Luykx et al., 2007, student responses are from p. 910). Clearly, each of these responses was influenced by the culture and experiences the children had in their home environments.
- Linguacultural influences: In their analysis of the students' written responses to assessment questions, the researchers found "confusion around discursive conventions for the

interpretation and production of scientific texts" (Luykx et al., 2007, p. 911). For example, a mathematical question stated:

> Your parents tell you to be home at 6:00 p.m. for dinner. It is now 4:00 p.m. How much time do you have to get home for dinner? Show your work." Most students were able to figure out "2 hours" but were unclear about the follow-up statement: Show your work. Many students left the item blank, some drew pictures of dinner tables or themselves playing at their home with their parent figure instead of showing the expected mathematical operation (i.e., 4 + ? = 6). (Luykx et al., 2007, p. 914)

Overall, the findings of Luykx et al.'s (2007) study questioned whether culture-free tests are ever feasible. If culture is an inevitable factor in assessment, then an understanding of non-mainstream cultures should be integrated as part of the process that informs assessment development.

Summary. Discussion in the literature regarding the challenges of assessing ELLs in science classes focuses on four main topics as outlined in Figure 7: (1) when the language of assessment is not the primary language of the test taker, limited content knowledge can be masked by limited language proficiency; (2) scientific terms have meanings that are different from everyday words and, while students may know the everyday definitions of terms, they do not necessarily know the meanings of the terms as intended in the context of science; (3) when dealing with an increasingly diverse and heterogeneous body of students, it is important to recognize the intrinsic role of culture in the way test questions are interpreted; and (4) science assessments inevitably contain culturally- and linguistically-implicit knowledge that is not accessible to all students because science does not operate in a cultural vacuum. Overall, the literature emphasizes that the failure to acknowledge these limitations threatens the integrity of the tools used to measure the fair achievement and competencies of all students.

Measurement Error

The literature about assessing ELLs raises concerns about the validity of the standardized tests that are widely utilized to make long-lasting decisions about students' abilities and educational futures. Factors that threaten the validity of commonly used assessments contaminate the analyses and inferences that can be made from the test results for this population of students. The fair assessment of students calls for an examination of factors that contribute to the measurement error of tests. The second section in this literature review focuses on issues of measurement error involved in the assessment of ELLs. In this section, I discuss literature that pertains to measurement error that threatens the test validity of standardized exams for ELLs, including issues related to language and dialect, the use of norming groups in test development, and the issue of construct irrelevance variance.

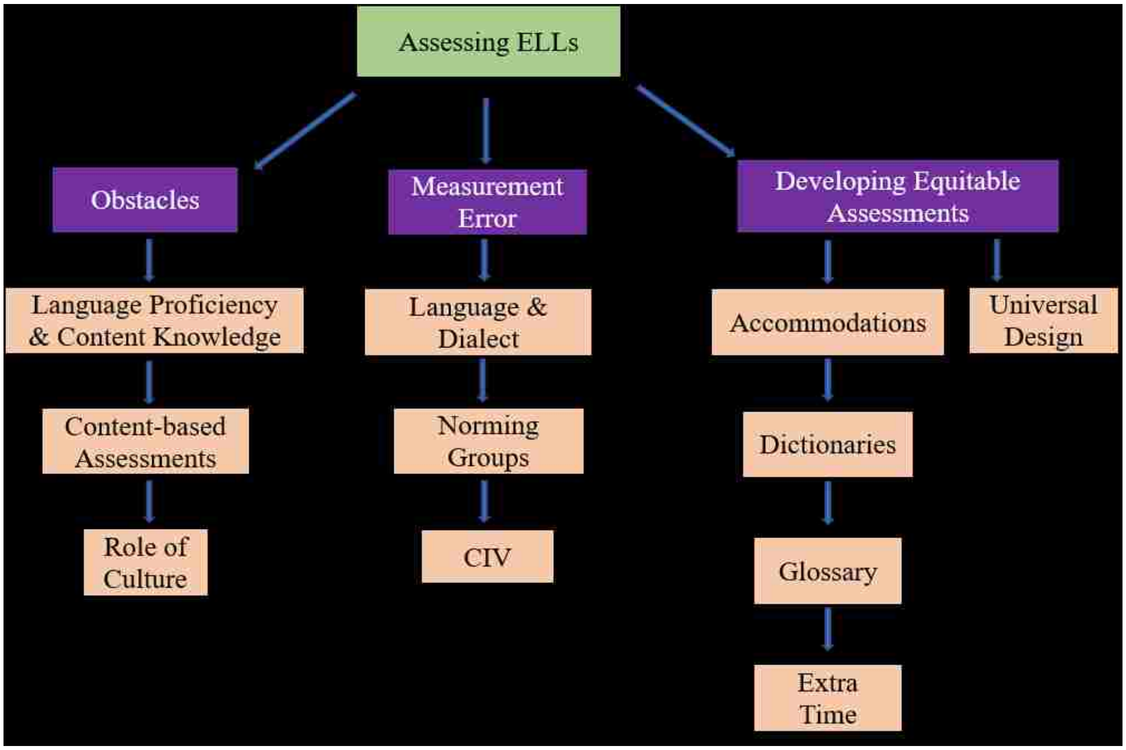


Figure 7. An overview of the "Measurement Error" section.

Factors contributing to measurement error. Discussion in the literature about assessing ELL students has been directed toward issues of test design, particularly issues that contribute to measurement error in scoring. Measurement error is defined as the observed error that occurs when there is a difference between a measured value and a true value. The main source of measurement error for ELL students stems from proficiency in the language in which tests are administered (Abedi, 2002). There are concerns that current approaches to test design do not account for language proficiency in evaluating a heterogeneous student population. Indeed, current test development methods only loosely address the role of how cognitive processes are shaped by language (Lee & Fradd, 1998).

Language and dialect. Solano-Flores and Li's (2009) study examined language and dialect as sources of measurement error that could threaten the validity of the National

Assessment of Educational Progress (NAEP) mathematics test for Grades 4 and 5. The study focused on the language of test administration (Spanish or English) and the dialect of Spanish on the test (local and standard dialects of Spanish) using test data collected from native Spanish-speaking ELL students. Fourth and fifth grade ELL students were given the same set of NAEP test items in either two languages or two dialects in order to examine score variation that was due to main and interaction effects of student, item, rater, and language. The analysis showed a significant score variation due to the interaction between student, item, and language, which suggested that each item posed linguistic challenges in each language and that each student had a unique set of strengths and weaknesses in each language. Similar results were found across dialects (Solano-Flores & Li, 2009). The results of the study highlight that measurement error in testing ELL students could result from assuming linguistic homogeneity in the population.

Norming groups. Another issue that contributes to measurement error is the student population for which a test is normed (Abedi et al., 2004). The commonly-accepted assumption that tests have been standardized across student populations is strained when considering that the norming group of test takers is mainly mainstream students. Generally, norming groups selected for test standardization purposes are not representative of a diverse student body and do not include ELLs. For example, the SAT9 included only 1.8% ELLs in its norming population even though this test is actively conducted in states, such as California, where ELLs represent 25% of test takers (Solorzano, 2008). The fact that many tests are normed with mainstream students implies that most test items are built on prior knowledge and learning experiences of the dominant group of students, not on those of linguistic minority students. Cummins et al. (1988) argues this point further as a threat to test validity:

> To the extent that their culturally-conditioned learning experiences differ from those of the majority group, minority children have less opportunity to learn the test content than majority children. The construct validity of the IQ test as a measure of previous learning [for minority children] automatically disappears since their previous learning experiences have not been adequately sampled. (Cummins et al., 1988, p. 267)

When tests that are normed for native English speakers are taken by students with a developing language proficiency, the resulting scores may not reflect an accurate measure of their ability. Instead, the test scores from such tests partially reflect English proficiency instead of content knowledge, competencies, and/or aptitude.

Construct-irrelevant variance. One of the central problems with applying standardized test scores to ELL students' performance is the failure to account for the language factor. Abedi (2007) contends that "there is no evidence to suggest that [ELLs] have less ability to learn content knowledge than NES students. Therefore, nuisance variables such as linguistic and cultural biases may mainly be responsible for such performance gaps" (p. 11). Consequently, measuring an extraneous construct, which is not related to the test's intended construct, causes a biased score distribution and is defined as construct irrelevant variance (CIV).

CIV is often described as the error variance arising from systematic error on test scores. Haladyna and Downing (2004) explained CIV using a linear model: "$y = t + e_r + e_s$, where y is the observed test score, t is the true score, e_r is random error and e_s is systematic error due to CIV" (p. 18). The extent to which CIV causes ELL students' scores on exams to deviate from those of NES students depends on the attributes of the test as well as those of the test takers. The quality of the test item and students' verbal abilities—which include reading, writing, speaking and listening—are among the salient factors influencing CIV for ELLs. Downing's (2002) work showed that poorly-crafted test items on a locally generated exam particularly impacted low-scoring students more than high-scoring students. Vocabulary terms used on test items have a significant impact on the CIV for students with limited English proficiency (Abedi et al., 2000). Time limits of a test also contribute to CIV because ELLs are often slower readers and require more time for comprehension (Fitzgerald, 1995).

Developing Equitable Assessments

In order to make assessments more accurate, meaningful, and equitable for ELL students, research suggests that additional measures be taken during test development (Solorzano, 2008). These include rectifying linguistically complex test items, including a diverse group of students in norming groups, and refining the test items to remove any language and cultural biases. However, virtually no such efforts have been made formally in large-scale testing procedures. This is, in part, due to the shortage of research that investigates assessment issues for ELLs. Kopriva (2002) highlights the need for this type of research:

> Undertaking studies that seek to understand what elements in the tests provide barriers for specific students, for what reasons, and what can be done to alleviate these barriers, will provide important construct-validity evidence. (Kopriva, 2002, p. 106)

This section focuses on literature that discusses the development of equitable assessment practices as outlined in Figure 8. Two main approaches have been suggested to make assessments more equitable: test accommodations and universal design. The following section includes a discussion of each of these approaches, starting with test accommodations.

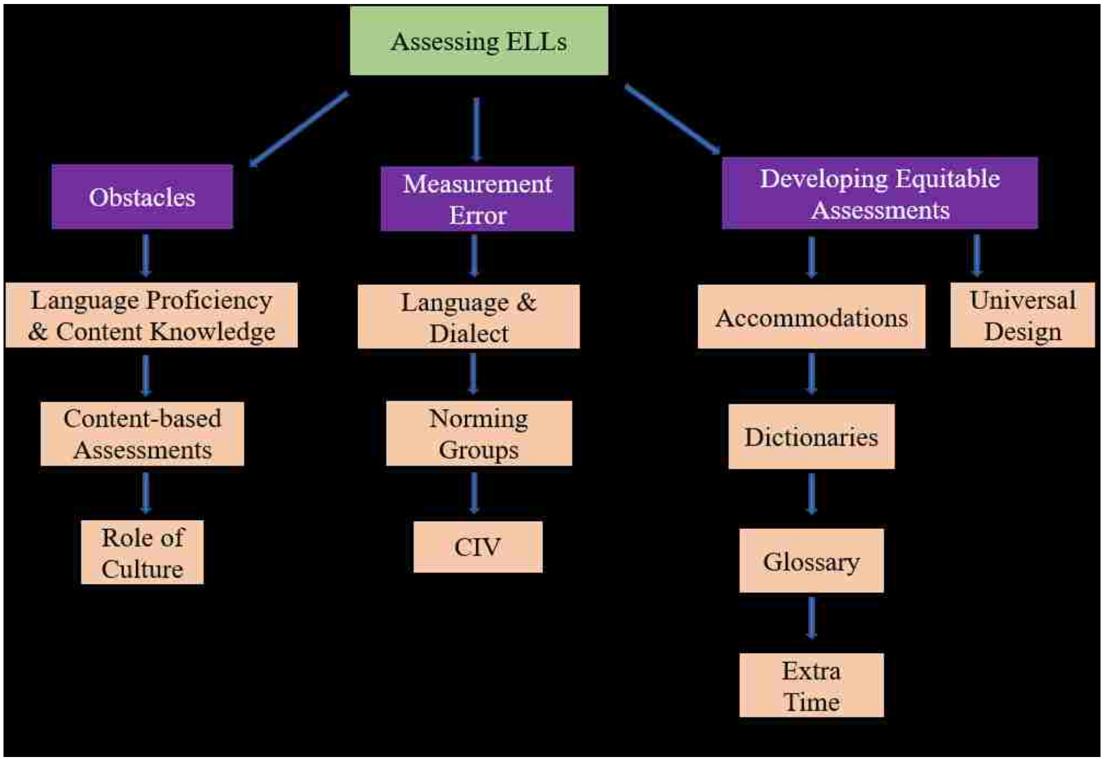


Figure 8. An overview of the subcategories under "Accommodations."

Test accommodations. Historically, the use of test accommodations for ELLs has been suggested as a way to improve test validity and accountability for ELLs by minimizing the construct that is not being measured. Rather than changing the test itself, proponents of test accommodations contend that ELLs should be provided with additional tools during the test to reduce impact of the language factor on their performance.

There are many types of test accommodations that can be used to aide ELL students during tests. Some of them include changing the environment in which students take the test (i.e., assigning ELLs to take the test outside of the classroom), increasing the amount of time allocated (i.e., providing more time to ELLs to complete the test), and allowing the use of additional equipment or materials during the test (e.g., dictionaries, translating tools, etc.). To evaluate the efficiency of accommodations for ELLs participating in large-scale assessments, Kieffer et al.

(2009) conducted a meta-analysis of seven different types of test accommodations that were empirically evaluated: (1) simplified English; (2) English dictionary or glossary; (3) bilingual dictionary or glossary; (4) tests in native languages; (5) dual language test booklets; (6) dual language questions for English passages; and (7) extra time. Their analysis reviewed 11 studies that included 23,999 participants (17,445 NES students and 6,554 ELLs) in order to compare the academic achievement test scores of ELLs with or without the accommodation with those of NES students. Their meta-analysis also evaluated the effectiveness of the accommodations in improving ELLs' performance, as well as the validity of these accommodations. The results indicated that only the use of English dictionaries or glossaries had a statistically significant impact on ELLs' test performance: there was a 10-25% reduction in performance gap between ELL and NES students when ELLs were allowed to use an English dictionary or glossary (Kieffer et al., 2009). The results from Kieffer et al.'s (2009) work suggest that test accommodations (with the potential exception of dictionaries) are mainly ineffectual in significantly reducing the performance gap between ELL and NES students on large-scale assessments.

Dictionaries. Other studies that examined the use of published dictionaries as a test accommodation provided to ELLs did not find similar results. Allowing the use of commercially-published dictionaries was thought of as a way to add language support to ELLs, as students would be able to look up the meanings of unfamiliar terms. Abedi et al. (2005) investigated the use of published dictionaries and bilingual dictionaries for 611 students in Grades 4 and 8 across several schools in the U.S with ELL and NES students on science tests. They reported that published dictionaries were not useful and were problematic to implement because certain definitions did not carry the appropriate meanings in the context of the test item. Mainly, these researchers concluded that commercially-available dictionaries and bilingual dictionaries provide definitions that are often broader, not content-specific and with varying levels of difficult vocabulary. In other words, commercially-available dictionaries provided definitions in the context of everyday talk and not in terms of contextualized science meanings.

Customized glossary and dictionary. Given the results presented above, providing students with a glossary of relevant definitions of unfamiliar words in the test booklet could be considered more effective than providing students with commercially-published dictionaries. Abedi et al. (2000) found that when provided extra time, both ELLs and NES students performed significantly higher on updated versions of math tests that included embedded definitions. Another study examined the use of three different types of accommodations among 422 8^{th} grade students' performance on a national science test. The test format included one test booklet with an English glossary, one with words translated in Spanish placed in the margins, and a test booklet with a customized English dictionary at the end of the booklet that included only the words that appeared in the test items. ELL students scored highest on the customized dictionary accommodation in the study, but there was not a significant increase in scores for NES students using the same accommodation (Abedi et al., 2000). Based on these studies, the use of customized dictionaries as an accommodation in the test itself seems to yield effective results for ELL students.

Extra time. Additional time is the most common type of test accommodation provided to ELLs in K-6 systems; however, its effectiveness is uncertain. Abedi et al.'s (2000) study investigated the effects of different accommodation strategies employed during a mathematics test administered to 946 8^{th} grade students, which included both ELLs and NES students. The accommodations provided included (1) a glossary, (2) extra time, and (3) a glossary plus extra time. Their findings indicated that, although providing students with a glossary and extra time resulted in a modest increase in scores, students' scores were most improved in the accommodation that included a (3) a glossary with extra time. Miller et al. (1999) investigated the effect of providing extra time, translating instructions, and/or providing a bilingual dictionary on ELLs' performance on a Grade 11 mathematics exam. They found that ELL students' scores were, overall, the highest under standard testing conditions (without accommodations) and declined under the condition of extra time. The findings from these two studies imply that offering additional time on exams does not necessarily help ELL students during test taking at all grade levels; however, when extra time is given in conjunction with the use of another accommodation (such as a customized dictionary), extra time could be useful.

Limitations of test accommodations. The idea of using test accommodations often raises controversial questions in the field such as: Who should receive accommodations? Who is eligible and who should decide eligibility? What type of accommodation is appropriate for each student? Is a one-accommodation-fits-all approach feasible for linguistic minority students? Even if using established criteria for eligibility, the literature suggests that there are legal and ethical issues (e.g., discrimination, equal opportunity, and grade pollution) associated with fairness about giving some students accommodations and excluding others, which has raised concerns among parents and administrators (Abedi et al., 2004).

One major limitation of using test accommodations is that the ELL student population varies tremendously in linguistic backgrounds. Consequently, their language proficiencies are diverse in English. Therefore, while a particular accommodation may be useful for one student, it may not be equally beneficial to another. Research strongly suggests that in order to make test accommodations more effective, accommodations should be developed individually for each student based on her or his language proficiency (Solano-Flores & Li, 2009). Unfortunately, this task is administratively cumbersome, costly, and unlikely to be implemented in undergraduate science courses at the postsecondary levels, the area of focus for the current study.

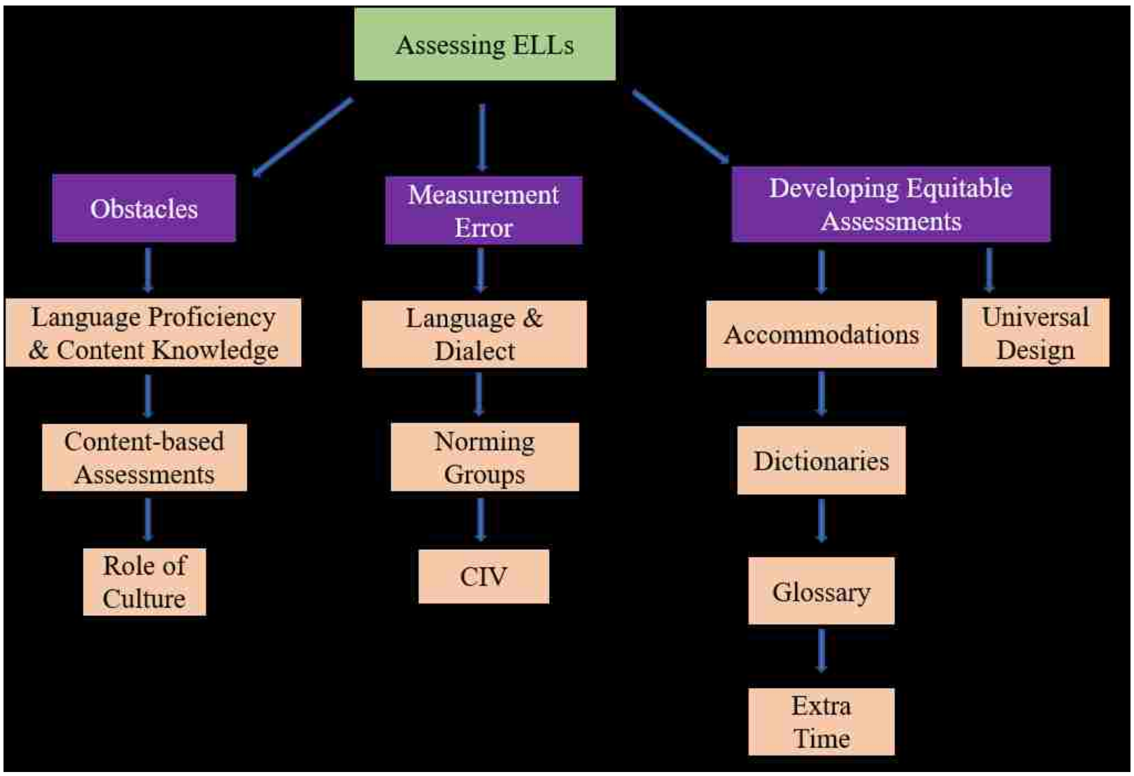


Figure 9. An overview of the “Universal Design” subcategory.

Universal design. Universal design (outlined in Figure 9) offers an alternate approach in making assessments equitable. The main idea underlying the universal design approach for assessments is that it transforms the test in a way that increases accessibility and equity for all students. Originating from the field of architecture, the goal of universally designed tests is to develop assessments that are accessible for the widest range of participants without threatening validity. In order to clarify tests and reduce the complexity added by irrelevant constructs, the following strategies are recommended in literature (Thompson et al., 2002):

- Reduce excessive length of sentences.
- Use common words and avoid unusual words (e.g., use “use” instead of “utilize”).
- Avoid passive voice and complex sentence structures.
- Avoid ambiguous words that may have more than one meaning.
- Avoid irregularly spelled words.
- Avoid proper names.
- Avoid multiple names for the same concept.
- Use well-designed graphic arrangements and headings to relay importance of information.
- Give clear and obvious signals to indicate separate questions.
- Place illustrations and/or schematics that contain information being assessed directly next to the test item.

The tenets of universal design promote the notion that language simplification of the test does not necessarily mean that the level of content knowledge being assessed has to be altered. Instead, the emphasis is on providing scaffolded language support that helps all students, including ELLs, better understand the questions within their linguistic boundaries. The benefit of universal design for the instructor is that it helps them better evaluate an ELL student’s content knowledge by decreasing the cognitive load of linguistic complexity. Pappamihiel and Mihai (2006) offered an example (shown below) that illustrates an application of such linguistic modifications on a middle school standardized assessment test in the mathematics section.

Original question:

An engineer is designing a metal gasket for a spacecraft. The gasket has the shape of a cylinder with a cylindrical hole through the center. The diameter of the gasket is 9 centimeters, and its height is 4 centimeters. The diameter of the hole is 3 centimeters. What is the volume of metal, in cubic centimeters, that is required to make the gasket? (Pappamihiel & Mihai, 2006, p.37)

Because of the level of unfamiliar vocabulary in the question and sentence structure, the question most likely reads to an ELL student like this:

An _________ is designing a metal ______ for a ________. The ______ has the shape of a cylinder with a __________ hole through the center. The diameter of the _______ is 9 centimeters, and its height is 4 centimeters. The diameter of the hole is 3 centimeters. What is the volume of ______, in cubic centimeters, that is required to make the _______? (Pappamihiel & Mihai, 2006, p.37)

In order to alleviate the complicated form of the question, a well-labeled diagram could be added without compromising the content knowledge targeted (Figure 10) and the question could be reworded as shown below.

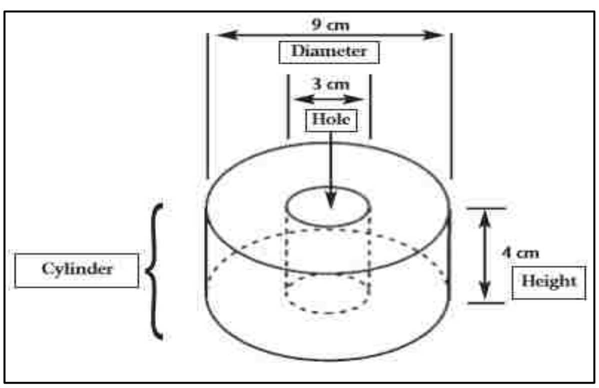


Figure 10. Modification of a math test question by adding a diagram (Pappamihiel & Mihai, 2006, p. 37).

Modified question:

A cylinder has a diameter of 9 centimeters, and its height is 4 centimeters. There is a hole in the middle of the cylinder. The diameter of the hole is 3 centimeters. You want to fill the cylinder with water. What is the volume of water that is required to fill the cylinder in cubic centimeters (cm^3)? (Pappamihiel & Mihai, 2006, p.37)

Universal design encourages the use of diagrams on science and math test questions to alleviate an overload of textual information. Martinello's (2009) study underscored the usefulness of including schematic representations in math word problems for ELL students. The study investigated the linguistic complexity of word problems as a source of differential item functioning (DIF) for ELLs on a 4^{th} grade math test. DIF contributes to measurement error, which occurs when people from different groups with the same level of ability or skills have a different probability of getting an answer correct on a test. The project examined 3,179 ELL and 65,660 NES students' scores from the statewide implemented standardized math test. A smaller focal group of 24 ELL students were selected for think-aloud interviews about the meaning-making processes they employed as they read the math word problems. The study discovered that multiclausal complex structures, unfamiliar vocabulary, references to mainstream American culture and text layout were the main hindrances for text comprehension. Findings indicated that the greater the test questions' grammatical and lexical complexity, the greater the difficulty estimates differences favoring NES students over ELL students. However, as designated by the universal design approach, it was confirmed that adding schematic representations on revised test items seemed to mitigate the effect of language complexity and alleviate DIF.

More support for the application of the elements of universal design came from Abedi et al.'s (2005) study that found that simplifying test items seemed to improve test scores for ELLs in higher grade levels. Abedi et al. (2005) studied the results of a standardized science test for Grades 4 and 8 ELL and NES students. The test was administered under four conditions: (1) no accommodation, (2) English dictionary, (3) bilingual dictionary, and (4) linguistic modification of test items. The results differed for each Grade level. English dictionaries worked better at Grade 4; however, linguistic modification of science test items worked best in Grade 8. This implies that, at higher grade levels such as in undergraduate courses, linguistic simplification tends to be more effective as content assessments include more higher-level academic language.

Summary. The discussion in the literature regarding assessing ELLs paints a troublesome picture of current assessment practices. Research emphasizes that as exams are administered in English, it is essentially difficult to determine the degree to which ELLs' test performance reflects their knowledge about the concepts or their language proficiency. There are important sources of measurement error that impede an accurate reflection of students' scores including language/dialect, norming groups, and CIV. In an effort to strengthen the reliability of test scores, the use of accommodations during testing has been employed, including customized glossaries, dictionaries, and/or extra time. Although employing test accommodations is a common strategy in PK-6 levels, studies have questioned the effectiveness of this strategy and found it to be controversial in terms of fairness. An alternate approach called universal design indicates that test items should be simplified in order to benefit the widest possible range of students. Studies show positive evidence for universal design in reducing the performance gap between ELL and NES students.

Chemistry Assessments

Many of the aforementioned studies regarding the obstacles in assessing ELLs and developing equitable assessments were based in the context of math and science. Because the purpose of this study is to examine chemistry test questions at the undergraduate level, it is important to briefly review the relevant literature on the current practices of chemistry assessments in the undergraduate setting. In this section, studies about the current assessment practices followed in the university undergraduate setting are discussed first, followed by a discussion of the general trends in alternative chemistry assessments at this level.

Current Practices in Chemistry Assessments

Although the National Science Foundation (NSF) recommends that science achievement should be measured using multiple methods (National Research Council, 1996), a large-scale study that looked at 28,576 science faculty members' grading practices revealed that chemistry and physics faculty used fewer assessment types in their courses than biology faculty. The study included faculty from both public and private and 2- and 4-year institutions across the U.S. The dominant method was found to be paper-and-pencil, multiple-choice items across chemistry and physics. Findings suggest that less than half of chemistry and physics faculty used assessment types that offered students opportunities to express their ideas through alternative modes such as essay answers or term papers. Data from the study suggested that faculty may perceive time constraints, limited resources and large class sizes of undergraduate classes as challenges to utilizing alternative strategies of assessment (Goubeaud, 2010).

Currently, most exams conducted in general chemistry courses in the university setting are in the multiple choice question format (MCQ), which has been considered standard practice in the field. It has been a challenge to develop alternate instruments to accurately assess the chemical knowledge of students (Hartman & Lin, 2011). The MCQ format is favored over other types of questions because this format can be answered by a large class of students and can be graded quickly with minimal error. However, literature suggests that MCQs do not provide deeper insights into students' understanding of key concepts compared with other formats such as short-answer, worked-out questions, etc. In fact, many MCQs in chemistry are focused on calculations which could be solved algorithmically and without an understanding of underlying concepts (i.e., these questions tend to be "plug and chug") (Nakhleh, 1993).

Hartman and Lin's (2011) study provided evidence for the claim that the commonly-used MCQ format in general chemistry exams cultivate the assessment of problem-solving algorithms in lieu of core chemical paradigms. The study selected MCQs from a pool of common questions from a database at the United States Naval Academy and followed the responses of 900 students to questions about specific general chemistry topics. Data analyses found that the percentage of correct answers (PCA) does not correspond to how advanced the topic was, and 47 to 93% of students can choose the correct answer based on the format of the question and not the content. For example, although students were able to correctly choose a response for the calculation of K_b from K_a, they were unable to correctly answer a conceptual question that asked them to interpret what the value of K_b means in the context of base strength (Hartman & Lin, 2011). These findings highlighted the fact that there is a lack of correlation between a student's ability to solve a problem algorithmically and her/his conceptual understanding of the question. Because a majority of general chemistry test bank questions (from mainstream textbooks and commercial test banks) involve algorithmic questions, there are concerns in literature about whether students are being taught problem-solving computations in order to succeed on exams in lieu of fundamental paradigms that pertain to general chemistry (Hartman & Lin, 2011).

Trends in Alternative Chemistry Assessments

Research suggests that a shift from more objective, paper and pencil based tests to less objective, student-centered tests is beneficial for assessing students in chemistry (Hartman &

Lin, 2011; Lewis et al., 2010; Noble et al., 2012; Wygoda & Teague, 1995). A study that implemented a restructured chemistry curriculum in high school found that that this approach increased student test scores on achievement tests and college entrance exams. Performance-based chemistry assessments were implemented in a first-year high school chemistry course. These included various activities that gave the students the opportunity to demonstrate their content knowledge. In order to best accommodate the changes in assessment procedures, the amount of material covered was reduced by selecting the most important concepts and skills to emphasize during the course. Students were assessed regularly on their scientific communication skills through writing and public speaking activities; collaboration skills were assessed as students worked in cooperative groups. For the "final" assessment, the researchers required students to perform a culminating demonstration of a cross-cutting chemical concept. Results of this alternative method of student assessment revealed that integrating alternative assessment elements to the content encouraged creativity and ingenuity in teaching and learning chemistry (Wygoda & Teague, 1995).

Another study that examined assessments in chemistry similarly found that shifting to a nontraditional form of assessment afforded students the opportunity to better demonstrate their knowledge (Lewis et al., 2010). Freshmen college-level science courses typically use traditional, multiple-choice questions or short-answer questions with only one correct answer (Lewis et al., 2010). This approach is considered a teacher-centered approach as the teacher chooses what students should know and then tests them on it. This approach provides only a partial picture of a students' knowledge in a course. The Lewis et al. (2010) study introduced an alternative assessment method in freshmen chemistry courses called Creative Exercises (CEs), in which students were given a statement or a prompt (for example, $H_{2(g)} + Cl_{2(g)} \rightarrow 2HCl_{(g)}$, on the topic of covalent bonding and electronegativity) and asked to write down as many distinct, correct, and relevant facts about the prompt as they can. This way, students were provided the opportunity to present their knowledge and were rewarded for the retention of previously presented concepts (Lewis et al., 2010).

Lewis et al. discovered that CEs were relatively easy to design and to grade, taking about an hour to grade CEs for a class of 70 students. CEs were especially insightful to instructors as they allowed the instructors to become aware of misconceptions and inappropriate conceptual connections that students had made. Additionally, through this method, students were able to show interconnectedness of chemistry topics through the course, which indicated a deeper understanding of concepts rather than remembering unrelated facts (Lewis et al., 2010).

Despite support in literature for alternate assessments in chemistry, current practices in general chemistry courses reveal that MCQs remain the preferred method of testing for undergraduate students.

Assessment of ELLs in Science

There is a lack of research that focuses on how undergraduate students perceive the characteristics of general chemistry assessments. Because an important goal of this study is to examine general chemistry test items from the perspectives of students, it was important to review any literature related to this topic. Accordingly, this section focuses on studies about student perceptions of science test items.

Siegel's (2007) study investigated the effects of original versus modified classroom assessments items. The original life science assessment items were selected by the teachers of two seventh-grade science classes whose students were participating in the study. The modified items were revised per the guidelines of the equity framework (discussed in detail in Chapter 3) to add linguistic support and prompts. The study quantitatively measured students' performance through a pre-test/post-test design. The pretest contained all original questions, and the post-test contained all modified questions. The research questions were aimed at understanding which modifications were needed to develop more equitable assessments, and how effective the modified items were for advanced ELL students compared to NES students. (Siegel, 2007).

The results of this study implied that science test items can be made more accessible for all students. Findings of this study indicated that both ELLs and NES students performed better on the posttest, which included modified items. Although ELL students' performance was significantly lower than that of the NES students on the pretest, the improvement in ELL students' scores on the posttest reduced this gap. The revisions based on the equity assessment framework yielded accessible items for all students (Siegel, 2007). A major premise of this study was that ELLs know more than they can demonstrate on content assessments. This premise was supported by the results, as refining the test items reduced the performance gap between ELL and NES students.

In order to gain a deeper understanding of students' interpretations of original versus science test items that had been modified according to the guidelines set forward in the Equity Framework for Classroom Assessment (EFCA, see Chapter 3), Siegel et al. (2014) conducted a qualitative investigation. The written assessment items were developed from disciplinary core ideas as outlined in the Next Generation Science Standards (NGSS) for life sciences, with topics including molecules, organism structures, ecosystems, energy, and dynamics. The original items were developed using typical discourse patterns of commonly used assessment questions. The goal of the EFCA revisions was to simplify the structure and language of the original test items to make them more accessible and approachable for ELL students. An example for how their item simplification was conducted as well as a list of all modifications applied is shown below:

Left: Zebra mussels clogging a pipe (1,2). *Right*: Smaller Zebra mussels attached to larger, native mussel (1,2)

~~Describe~~ What has been the impact of zebra mussel on ~~Missouri ecosystems~~ ecosystems in Missouri in terms of ~~competition and the mussel' effects on man made structures~~ the following: (3)

- Competition with other mussel species (4)
- Effects on man-made structures and equipment (dams, boats, etc.) (4)

Modifications:

1. Addition of visual support
2. Brief note describing each visual support.
3. Reduction of words in the item stem.
4. Replacing a long question with two simple questions. (Siegel et al., 2014, p. 686)

Data were collected in two forms: written data on students' notes as they solved the questions, and interview data from think-aloud protocols and post-assessment questions. The students found that both versions of the question were challenging in terms of content and struggled with talking through their reasoning; however, the modified versions provided prompts that better facilitated their understanding. The modified items increased students' comprehension of the questions, elicited responses, and helped students with organizing and visualizing their thoughts, which enabled them to perform better on the question than they did without the modifications (Siegel et al., 2014).

The two aforementioned studies highlight that revising science assessment questions with structural modifications, such as multiple prompts and graphic organizers, and visual tools, such as photographs and diagrams, enhances students' abilities to respond to those questions in a meaningful way (Siegel, 2007; Siegel et al., 2014).

Summary. The research literature about language acquisition and assessment indicates that more research is needed to design test questions that adequately reflect ELLs' content knowledge instead of their limited proficiencies in the English language. As outlined in the literature, the development of valid assessments for ELLs is not a straightforward process as it involves the consideration of key issues at stake, including academic language proficiency, culture, using norming groups that better represent ELLs during test development, and accounting for construct irrelevance variance. Educators have attempted to address these issues in two major ways: (1) offering students accommodations during tests, and (2) utilizing a global approach, called universal design, aimed at making tests accessible to all students. Siegel et al.'s (2007, 2014) work demonstrated that there is value in revising original classroom assessment items for all students. However, this work was focused on middle school ELL and NES students in life science classrooms. Additional work is needed to determine if the modifications suggested by the EFCA are equally valid and effective for the undergraduate chemistry student population, including both ELL and NES students.

Justification for Current Study

According to the President's Council of Advisors on Science and Technology (PCAST, 2012), one million additional more STEM graduates should be produced over the next decade to meet the demand of a STEM-ready workforce in the U.S. During a time of economic recession in the U.S., the unemployed outnumbered jobs vacancies 3.6 to one for non-STEM jobs; however, for STEM occupations, there were twice as many jobs per one qualified person (Change the Equation, 2010). Racial and ethnic minority students represent a largely untapped STEM talent pool in the U.S. Minority students are rapidly increasing in the national education system; however, low rates of success among minority students in STEM education persist (Museus et al., 2011). Although linguistic minority students are an important subgroup of this underrepresented population, they are often overlooked. Reports indicate that the ELL population is rapidly growing in the education system (ECS, 2014); however, there is little research that examines their specific needs to succeed in STEM courses.

Research indicates that there is significant attrition among STEM majors, especially in the first two years of college, during which many students enrolled in introductory science courses choose to switch majors (Rask, 2010; Seymour & Hewitt, 1994; Tobias, 1990). The way students experience assessments in their undergraduate coursework could play a critical role in their decision to continue in STEM majors. Research indicates most science assessments do not sufficiently factor in the language and cultural backgrounds of students and fall short of adequately assessing students' content knowledge. However, there is a short supply of research studies focused on gaining a deeper understanding of how undergraduate students perceive their course assessments. There are even fewer research studies focused on ELLs' perceptions of science assessment items. Furthermore, there are currently no research studies that are aimed at studying the perceptions of undergraduate ELL and NES students about general chemistry assessments.

One major goal of this study was to revise original general chemistry assessment items based on the theoretical principles of the equity framework for classroom assessments (see Chapter 3), with the goal of making items more accessible and clearer for all students to understand and answer. Unlike most other studies in the field that have used quantitative approaches to evaluate student scores to validate an assessment method, this study was designed to gain insights through the experiences of students who are actively attempting to think through each assessment item presented to them. This included students' perceptions of typical chemistry assessment questions as well as their perceptions of assessment items that were modified according to the principles of the equity framework for classroom assessments. Specifically, I was interested in knowing which features of the questions students find beneficial and/or confusing for interpreting and answering each question as intended. This research study addresses a gap in literature by explaining how NES and ELL undergraduate students perceive general chemistry assessment items—both commonly-used assessment items and items that have been revised according to the equity framework for classroom assessment. The results of the current study will benefit instructors, administrators, and other entities with a vested interest in creating equitable assessments for all students pursuing STEM fields.

CHAPTER 3 THEORETICAL FRAMEWORK

Overview

The Equity Framework for Classroom Assessments (EFCA) is a relatively new conceptual framework. The goal of this framework is to modify test items in order to make testing more equitable without reducing the difficulty of the content being assessed. In this chapter, I first discuss the development of the framework, followed by a description of the framework, including its main principles. The later sections include a discussion of data collection and data analysis in studies informed by this framework. This chapter also includes an explanation of the framework's limitations and a justification for its use in the current study. The final part on this chapter contains a definition of each item modification derived from the framework, which modifications were applied to the general chemistry assessment items used in this study.

The Development of the Equity Framework

The equity framework for classroom assessments has its underpinnings in sociocultural perspectives of learning, which profess that individuals are continually shaping and are being shaped by their social environments. This view is rooted in Vygotskian beliefs that human consciousness is not an internal property of the subject or an interior element; rather, it is the product of the individual's interaction with the social world and carries a dialectical and mediated character (Leont'ev, 1978). This implies that the individual engages in discourse with others to establish truth through negotiated reasoning. Vygotsky theorized that language is the key in mediating knowledge. From this position, an individual cannot be a fully internal being isolated from external experiences; she/he, then, develops on the basis of her/his cultural and historical resources, including language (Peim, 2009).

A Brief History of the Equity Framework

The EFCA started as a set of assessment principles for science teachers of ELLs (Siegel, 2007). At the time of the development of the EFCA, most research on assessing ELLs focused on accommodation strategies that included using a translating tool, dictionary, glossary, or providing extended time limits on large-scale assessments. The researchers that developed the EFCA noticed that, when different accommodation strategies were tested (extra time, glossary, glossary plus extra time, and linguistically modified test items), the only strategy that reduced the performance gap between ELL and NES students was linguistically modified items (Abedi et al., 2000). Therefore, the researchers designed the EFCA as a new approach to linguistic modification that can be implemented in diverse classrooms in an effort to improve assessments (Siegel, 2007).

Sociocultural perspectives contributed to the development of the EFCA. Here, I briefly discuss the tenets of the sociocultural perspectives which apply to the EFCA. The sociocultural point of view of learning contends that in order to understand how students learn, it is critical to see students as part of a community outside of the classroom, with their own cultural and sociological interactions that influence their sense-making processes. The need for sociocultural perspectives on science and science education became more evident as researchers challenged the view that science takes place in a closed system, disconnected from social institutions, cultural beliefs, and values (Haraway, 1989, 1991, 1999; Latour, 1987; Shapin & Schaffer, 1985). The premise that science education must be examined as a human activity in the context of the dominant culture and political issues of the time has become increasingly accepted (Lemke, 2002). According to Lemke (2002), learning and doing science is “primarily socially learned cultural traditions of what kinds of discourses and representations are useful and how to use them, far more than whatever brain mechanisms may be active while we are doing so" (Lemke, 2001, p. 298). In the EFCA, these ideas have been extended to consider of how minority students understand written scientific discourse. The culminating premise of these ideas is reflected in the main principles of the EFCA, which are discussed in the next section.

Description of the Equity Framework

Assessing the content knowledge of an increasingly diverse body of students is understandably challenging. Research from Okhee Lee (2004) suggests that a majority of teachers are unaware of ways to teach (and test) ELLs the English language in the context of science. For instructors who may not have the tools and/or experience in evaluating ELLs, designing test items that are both fair and valid can be a daunting task. Rather than making the tests easier for ELLs and/or giving them additional accommodations (e.g., extra time, glossary, translator, etc.), the equity framework designates five principles that reduce linguistic complexity and make test items more accessible for all students. In this section, I discuss the five main principles of the equity framework for classroom assessments and how they are applied to the modification of test items to make them more accessible:

1. Assessments should match the learning and instructional goals.
2. Assessments should be linguistically and culturally comprehensible.
3. Assessments should challenge students to think about difficult ideas.
4. Assessment should elicit student understanding.
5. Assessment should scaffold the use of language and support learning. (Siegel et al., 2008, p. 44)

Assessments Should Match the Learning and Instructional Goals

"The assessment *is* the curriculum, as far as the students are concerned. They will learn what they think they will be assessed on, not what is in the curriculum, or even on what has been 'covered' in class" (Biggs, 2003, p. 3). Because of this, an important goal for classroom assessments should be to match the learning tasks and/or activities done in the classroom. Biggs (2003) uses the term *constructive alignment* to describe this notion. In constructive alignment, students are making meaning through relevant learning tasks in class, and assessments are designed to correlate with those tasks. For example, if the assessment in the course involves a seminar talk or presentations, then instruction should emphasize communication skills of the discipline and work on activities that expound upon these skills. This model drives teachers to think about what they want their students to learn and explicitly work to clarify learning outcomes (Biggs, 2003).

Constructive alignment is particularly advantageous for ELL students because their classroom tasks mirror their assessment tasks. Classroom activities play an important role in teaching ELLs how and when to use academic language skills in context. For example, if an upcoming course exam includes extended essay responses, classroom learning tasks can be designed to include sample questions and sample essay responses. Biggs (2003) contends that what is assessed and how it is graded sends a message to students about the type of knowledge and skills that are most valued in a discipline. Therefore, if an instructor is trying to downgrade the language factor (inherent linguistic components that can pose as barriers to comprehension) in the test items, she/he may choose to include assessments that allow students to holistically demonstrate their knowledge of a concept by allowing students to include non-verbal components with their written responses (e.g., illustrations, drawing, showing an experiment, etc.). Per Biggs (2003), this type of adjustment to assessments would convey the message to students that understanding of how concepts work is important and/or more important than verbatim responses or rote memorization of facts.

Based on the concept of constructive alignment, Siegel (2007) suggests that modified written items should match the concepts, scientific goals, and the discourse of the original written item. This principle ensures that revised items maintain the conceptual rigor of the original item. To make certain that revised items are more accessible than their original counterparts, the language of the revised items must also be consistent with the language of instruction (Center for Research on Evaluation Standards & Student Testing [CRESST], 2001). For example, if the word *trial* is used during teaching, then the word *experiment* should not be used on the assessment, even though the two words may have equivalent meanings to a native English Speaker. ELLs may not be familiar with synonyms of discipline-specific terms, and this unfamiliarity results in an increased reading time and/or unintended evaluation of academic language terms rather than understanding of a concept.

Assessments Should Be Linguistically and Culturally Comprehensible

Language is so much a part of teaching and assessment that educators seldom stop to acknowledge the innate role it plays in constructing meaning. Unless instructors are trying to teach ELLs and/or students with language learning disabilities, the linguistic skills and issues that can impact student success go largely unrecognized. Mainstream instructors may not be aware of many linguistic demands that assessments pose on students. The second principle of the equity framework addresses these linguistic demands and aims to make assessments more accessible to ELLs.

It is important for instructors to recognize common language barriers that are present in test items. Science questions are often decontextualized, and contain unfamiliar words, complex sentences, and grammatical characteristics that are difficult to follow. Science test items tend to be decontextualized in nature. As a consequence, students may require added support to understand the questions as intended by the test maker. Because outside support is not typically available during testing, ELLs may not be able to get clarification on the intended meaning of the test question. In this case, they are forced to bring their own extra-linguistic context, which is influenced by their personal backgrounds, to the test item.

Unfamiliar vocabulary is another issue that is known to add to linguistic barriers of assessments. The terms that cause linguistic barriers are not always scientific terms. Instead, words that have different meanings in everyday language versus technical language are more difficult to follow. For example, *likely* (expresses something that is probable, not the act of liking something), *respectively* (expresses the order things mentioned, not having respect as a virtue), *significant* (statistically different, not important in common terms), are used to relay scientific reasoning (Abedi et al., 2005).

Sentence and text complexity also pose barriers to comprehension. For example, the following question contains double negatives, "Under what conditions is it not impossible to float a lead canoe?" Other problematic grammatical characteristics within sentences include long phrases in questions, compound sentences, logical connectors, long noun phrases, relative clauses (starting with "who/whom," "that," "which"), lengthy problem statements, passive voice, and poor cohesion across paragraphs (e.g., lack of transition words such as "then" or "next.") (Trumbull & Solano-Flores, 2011).

The equity framework denotes that because language is crucial in conveying meaning, test items must be revised in a way that reduces linguistic barriers. One important goal is to minimize the unnecessary linguistic complexity of assessments so that ELLs are not forced to spend extra time reading the test item compared to NES students. Based on this principle, the following measures should be taken:

- Sentences should be shortened and simplified.
- Ideas should be bulleted to reduce reading time.
- Pictures and/or other illustrations should be added in place of words when possible.

The following example shows how this principle can be applied:

> Original item: The scientists interviewed the patients to find out whether their coughs were as frequent and as serious. They also asked the patients if they had any new health problems while taking the medicine.
>
> Revised item: At the end of one week, the scientists asked the patients:

- Is your cough better, the same or worse?
- Do you have any side effects, such as dizziness or upset stomach? (Siegel et al., 2008, p. 9)

Because culture is a major component of students' prior knowledge, the equity framework contends that student learning should be assessed by factoring in the role of culture (Fong & Siegel, 2005). Culture is defined as "subtle and invisible […] a collection of values, beliefs, and standards which influence how students think, feel, and behave in various social setting including classrooms" (Readence et al., 2004, p. 31). Assessment, then, not only evaluates students' cognitive capabilities, but also their sociocultural backgrounds. Research suggests that cognition and reasoning skills differ across cultures (Ji et al., 2004). This is especially the case on science assessments, where critical thinking and problem solving skills are probed.

> The way students interpret science items and respond to them may be more influenced by personal experience than formal school learning experience. Frequently, everyday life experiences seem to be what first comes to students' minds when they respond to science items. (Solano-Flores & Nelson-Barber, 2001, p. 559)

When designing assessments, it is important to be mindful of students' backgrounds and how they may hinder students from correctly interpreting test items. Linguistic and cultural assumptions underlying a test item may interfere with the intent of that assessment task and may not be applicable to non-mainstream groups of students, such as ELLs, low socioeconomic status students, immigrant students, etc. For example, a physics problem that uses a golf course as the context to solve a problem would be assuming that all students taking the test are familiar with golf courses. However, economically underprivileged inner city students who may not have seen a golf course would be placed at a disadvantage when attempting to answer such a problem (Siegel et al., 2008).

Creating linguistically- and culturally-accessible assessments is a difficult process. The equity framework emphasizes that removing bias altogether from test items is not possible. However, the goal is to reduce biases that are due to differences in race, culture, economics, gender, and language by modifying and refining assessment items where assessment biases might exist (Siegel et al., 2008).

Assessments Should Challenge Students to Think About Difficult Ideas

Another important recommendation that the equity framework makes is that assessments should challenge students to think about tough ideas. This notion urges instructors to create assessments that are rigorous in content. Unfortunately, watering down content and reducing the curriculum is done too often to accommodate ELL students, and this leads to denying students an equal opportunity to learn (Walqui, 2003). Instead, equitable assessments should remain challenging to promote intellectual growth and preparedness for more advanced concepts. The following techniques are suggested to help students deal with linguistic issues while the content level is maintained (Note: examples of modifications for each of the following have been included at the end of this chapter, under the heading "Item Modifications"):

- "Simplify vocabulary and grammar;
- Use scaffolding;
- Provide customized dictionary;
- Offer word bank for responses;
- Use interpreter;
- Read questions aloud" (Siegel et al., 2008, p. 9).

Additionally, Garcia and Pearson (1994) advocate for the use of written and materials-based performance assessments to assess the knowledge of ELL students. Performance assessments prompt students to take multiple steps in completing tasks that utilize the skills, knowledge and dispositions for a discipline. For example, an activity is presented to students which would involve results of an experiment. The students would have to think of multiple operations to determine how variables X and Y are related and discuss outcomes. The task would be scored holistically in terms of reasoning, problem solving, communication, and connections. In this manner, assessments are not necessarily made easier, but they are contextualized and operationalized to measure students' understanding.

Assessments Should Elicit Student Understanding

The fourth principle of the equity framework focuses on better regulating students' responses on test items to match the intended response. The premise of this principle originates from White and Gunstone's (1992) work, which suggests that effective assessments should probe students' understanding and help students express their ideas. If students' responses reflect misinterpretations of the prompt and/or skipping parts of the prompt, then those items must be revised to be more effective. Test items that are written very generally often lack the direction students need to correctly follow the prompt and come to the outcome intended by the question. The following is an example of how an item was changed to elicit thinking:

> Original prompt: Should Rita stop taking antibiotics or finish the treatment? Explain the advantages and disadvantages of stopping and of continuing the antibiotics.
>
> Revised prompt:
>
> What are some good and bad things about stopping the full course of antibiotics? What are some good and bad things about continuing to take the full course of antibiotics? Write your answers in the table below:

Good things		**Bad things**	
Stop taking antibiotics	Continue taking antibiotics	Stop taking antibiotics	Continue taking antibiotics

(Siegel et al., 2008, p. 45)

White and Gunstone (1992) describe general guidelines for designing assessments to stimulate thinking. The following strategies offer students multiple opportunities to express their learning:

- Questions should be based on expanding ideas taught, not just recall.
- Questions should demand students' thinking about fact-based answers.
- Assessments should ask questions that start with "how would…" or "what if…"
- Questions should provide stimuli (e.g., quotations, data tables, maps, or diagrams) for questions.
- Questions should provide a large concept, and ask for thoughts on unanswered questions (e.g., "when a cell dies, does their DNA die as well?").
- Questions should be written about X to test students' deeper understanding of Y. In assessments, write questions about X that would test students' deeper understanding of Y.

Assessments Should Scaffold the Use of Language and Support Learning

The equity framework states that assessments should provide scaffolds for ELL students to support language use during testing, which can enhance their comprehension of the question. The term scaffolding means to add support to a structure and then gradually remove it. Scaffolding is an established learning strategy that has been used for ELLs to enrich and amplify the learning of content without diminishing the level of difficulty (Walqui, 2003). Teachers use scaffolds to provide temporary support that is later removed during classroom activities. When applying scaffolds to chemistry test items, the goal would be to embed a series of linguistic scaffolds to offset the high cognitive load of reading, translating, interpreting, and solving a chemistry problem. Adding scaffolding in the test items may help students comprehend the question, think about the topic and respond to the question better (Siegel et al., 2008).

The following scaffolding approaches can be applied to assessments: modeling, contextualizing, sentence starters, and graphic organizers, each of which is discussed below. Modelling offers students a representation of learning that can be used as a standard to imitate. Although modeling is primarily done during instruction for students, it has direct benefits in assessment as well. When students are given clear examples of what is expected of them in terms of responses to test questions, they are more comfortable approaching the assessment tasks. A 10^{th} grade ELL student responded positively to this technique:

> In my chemistry class I can always do well because the teacher first demonstrates an experiment, and then we try a similar one. Then he asks us to write down the procedure and the conclusions in groups of two or four. I can do it. I can even use the new words because I know what they mean. (Walqui, 2000, p.94)

Research suggests that, in addition to modelling, adding sensory language to increase contextual information may reduce the demands of academic language. Because academic language is context-reduced and can be different from everyday language, embedding sensory language, pictures, prompts, and/or analogies can help ELLs better grasp the intended meaning of the test item (Walqui, 2003). For example, in order to describe the role of mitochondria in generating ATP, a teacher could use the word “furnace” or “powerhouse” to help them imagine a furnace generating heat energy and also show short clip of this cellular structure (Walqui, 2003).

Sentence starters and graphic organizers can also serve an important purpose by providing scaffolding for test items. Sentence starters can give students additional clues about where a question is leading them. For example, “these groups are similar/different because…”, “one has…,but the other does not,” and “when…it causes…” Inserting phrases such as these in the beginning of the question provides signal words and messages about the purpose of the question. Graphic organizers can also be crucial in reducing the language complexity of a question prompt by visually arranging relevant information for students. Concept maps, K-W-L charts (What I **K**now- What I **W**ant to know- What I **L**earned), T-charts, and Venn diagrams are a few example of graphic organizers that have been suggested to use in order to scaffold test items (Siegel et al., 2008). One example of a graphic organizer is shown below:

Matter can be classified as shown on the chart below.

Based on this classification, which type of matter is air?

[redacted]

Summary. The overarching goal of the EFCA is to make assessments more accessible for all students. Because students interpret language differently, the five philosophical principles of the EFCA provide foundational support for designing assessments that:

- are aligned with instruction,
- consider students’ language and cultural backgrounds,
- enable student to think critically about content,
- elicit student learning through assessments, and

- provide scaffolded cues to facilitate comprehension.

These five principles of the framework led to the development of a list of modifications that can be implemented to revise test items, each of which will be discussed in detail later in this chapter.

Methods & Analysis

As a relatively new framework, the EFCA has started to receive more attention in literature; and two recent studies—one using quantitative methods and one using qualitative methods—have directly applied the EFCA in research. Because the premise of EFCA is to transform assessment questions into more accessible items for ELL students, both studies focused how the students (ELL and NES students) interpreted and performed on the test items. The purpose of the quantitative study was to identify ways to improve written assessments for ELLs (Siegel, 2007). The qualitative study focused on the use of scaffolds in written classroom assessments through the perceptions of ELL and NES students (Siegel et al., 2014). In this section, I focus not on the individual studies, but on the data collection and analysis techniques used in both studies. This section is arranged in the following order: (1) typical participant pools, (2) item modifications, (3) typical methods of data collection, and (4) typical methods of data analysis used in EFCA studies.

Participants

In both of the studies that have been carried out with the EFCA as a guiding framework, the participant pool has included ELL students. Given that the goal of the EFCA is to make assessment items more equitable for ELLs, it is important that the voices and experiences of ELLs are a major focus of any study informed by the EFCA. Although the major focus of a study informed by the EFCA will be on ELL students, it is possible to include NES students in addition to the ELLs as participants. In theory, the modifications suggested by the EFCA should make assessment items more accessible by all students, so it may be important to include the voices of NES students, in addition to those of ELLs in a study informed by the EFCA (Siegel, 2007; Siegel et al., 2014).

Item Modification

Because an important goal of the EFCA is to revise and improve original assessment items based on the modifications indicated, the process of developing revised assessment items precedes data collection. Original science questions are identified from specific topics, which are drawn from course content. These original items are revised and refined with modifications such as reduced linguistic complexity, simpler sentence structures, reduction of non-essential information, etc. (Note: a list of explanations of each modification used in this study is provided below.)

The five principles of the EFCA discussed previously provide the rationale for specific item modifications that should be implemented to revise original test items. However, the original work does not propose operationalized definitions of each item modification utilized in the framework. To address this limitation, I have devised operational definitions of each type of item modification based on how each modification has been applied in previous EFCA studies and my own understanding of the principles of the EFCA. The operational definitions for each item modification are provided below, and in Appendix A.

Not all modifications described here will be necessarily applied to every original test item; some original test items may require more or fewer modifications than others depending on the nature of the item itself. To emphasize how each modification has been applied, only parts of a question are included as examples. To see the complete original test items and their completed modified versions that have been included in this study, please refer to Appendix B. To see the specific ways in which each question in this study was modified, please refer to Appendix C.

1. Linguistic simplification of vocabulary and syntax:
    a. Linguistic simplification of vocabulary: removing unnecessarily complex words and/or phrases and replacing them with simple terms that convey the same meaning
    b. Linguistic simplification of syntax: Replacing long sentences with embedded commas and/or semicolons with shorter and more direct statements that convey the same information in terms of content

    An example below shows how the first two statements of a general chemistry item have been modified.

Original	**Modified**
Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline.	Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars.

2. Replacement of sentences with lists:
    a. Modifying sentences that give more than one piece of information and/or adding a bulleted list to separate the pieces of information

    An example below shows how long statements from a part of a general chemistry item have been converted into a list.

Original	**Modified**
Calculate the theoretical yield of C_2H_5Cl when 125 g of C_2H_6 reacts with 255 g of Cl_2, assuming that C_2H_6 and Cl_2 react only to form C_2H_5Cl and HCl. Calculate the percent yield of C_2H_5Cl if the reaction produces 206 g C_2H_5Cl.	A. Using the balanced equation you wrote for Part A, find the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2. B. If you conducted the reaction described in Part B in the laboratory and only collected 206 g of C_2H_5Cl, what is the percent yield of C_2H_5Cl?

3. Reduction of nonessential information:

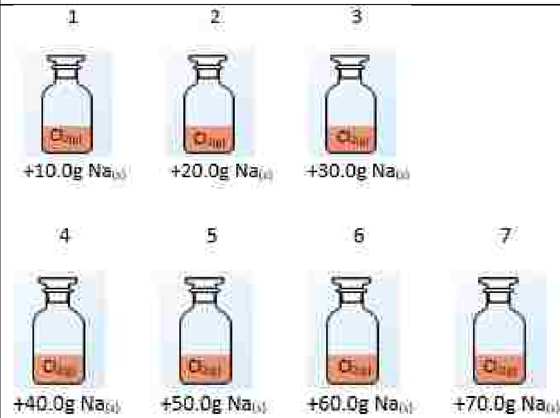
    a. Reducing the *number of words* in the item by removing unnecessary words that add to the overall reading time for students. This also includes removing extraneous words that are not necessary to understand and solve the problem, and removing content information that is irrelevant to solving the problem

An example below shows how statements from a general chemistry item have been modified by taking out unessential words such as *raw*, *material*, and *roasting* from the stem of the question.

Original	Modified
The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3.	Chromite, a chromium-iron ore, is a source of chromium and chromium compounds. By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4.

4. Addition of visual supports in the stem of item:
    a. Making the information in the item more visually accessible by formatting the question in a way that easily differentiates background information in the question stem (portion that gives background information) from pertinent information in the question. For example, adding paragraphs, line breaks, adding/removing space between background information, etc.
    b. Adding an illustration that encapsulates the information discussed in the item.

An example of a general chemistry item below shows how adding an illustration of jars in the beginning of the question strengthened the description of the problem.

Original	Modified
You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.	 There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows: Jar 1: 10.0 g $Na_{(s)}$ Jar 2: 20.0 g $Na_{(s)}$ Jar 3: 30.0 g $Na_{(s)}$ Jar 4: 40.0 g $Na_{(s)}$ Jar 5: 50.0 g $Na_{(s)}$ Jar 6: 60.0 g $Na_{(s)}$ Jar 7: 70.0 g $Na_{(s)}$

5. Division of data:

    a. Rearranging the order in which the information appears in the question so that it is logical and easy to follow.

    b. Adding a data table that contains critical information relevant to the problem.

An example below shows that this item was revised by letting the student know about the 92.0% yield before asking for the mass of sodium sulfide.

Original	**Modified**
How many grams of sodium sulfide are formed if 1.25 g of hydrogen sulfide is bubbled into a solution containing 2.00 g of sodium hydroxide, assuming that the sodium sulfide is made in a 92.0% yield?	You are told that this reaction gives you a 92.0% yield for sodium sulfide when you do the reaction in the laboratory. • How many grams of sodium sulfide will you make in the laboratory if you mix 1.25 g of hydrogen sulfide with 2.00 g of sodium hydroxide?

6. Alignment with the language of instruction:

    a. Monitoring the level of academic language used during instruction and matching the level of vocabulary in exam items. This would be outside the scope of my study because I will not be attending participants' classrooms. Therefore, I have not included an example of this modification.

7. Alignment of the language within the item:

    a. Ensuring that the tense, voice, and overall structure of the item are consistent.

An example below shows statements from a general chemistry item that reflect confusing tenses: mistake *is* made and we *used*. These statements were modified to be consistent in the overall voice and tense.

Original	Modified
What if a mistake is made and we used 2.37 grams of $Ca(OH)_2$ and 2.69 grams of H_3PO_4? How much $Ca_3(PO_4)_2$ will be obtained?	We made a mistake in our measurements. Instead of using the amount of calcium hydroxide you calculated in Part B, we used 2.37g of $Ca(OH)_2$ and 2.69g of H_3PO_4. How many grams of $Ca_3(PO_4)_2$ should be made under these conditions?

8. Use of bold type for emphasis:

   a. Highlighting an important phrase or word(s) in the item that is crucial for solving the problem.

   An example below shows statements from a general chemistry item that gives important information to students about setting up a balanced equation based on the reaction described. In this case, the compound formulas were added and darkened.

Original	Modified
Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen.	Methanol is made by mixing gaseous carbon monoxide (**CO**) and gaseous hydrogen (**H_2**).

9. Addition of graphic organizers in the prompts:

   a. Adding illustrations and/or representations in order to better describe the problem

An example below shows that chemical formulas for calcium hydroxide and calcium phosphate were added as a form of representations.

Original	Modified
How many grams of calcium hydroxide are needed to make 3.75 grams of calcium phosphate?	Calcium hydroxide *($Ca(OH)_2$)* and phosphoric acid *(H_3PO_4)* react to produce calcium phosphate *($Ca_3(PO_4)_2$)* and water.

10. Division of the prompts into smaller units:

a. Dividing the question prompt into smaller, more comprehensible units if the question prompt is more than three sentences.

An example below shows a general chemistry test item that was written in a visually unstructured manner. This item was restructured by breaking apart the prompt into smaller, more comprehensible units.

Original	Modified
How many grams of calcium hydroxide are needed to make 3.75 grams of calcium phosphate? What if a mistake was made and we used 2.37 grams of $Ca(OH)_2$ and 2.69 grams of H_3PO_4? Which compound is the limiting reagent? How much $Ca_3(PO_4)_2$ would be obtained? What if 2.37 grams of $Ca(OH)_2$ were used along with excess H_3PO_4 and only 2.98 grams of $Ca_3(PO_4)_2$ were obtained (instead of the theoretical yield)? What would be the percent yield?	Calcium hydroxide ($Ca(OH)_2$) and phosphoric acid (H_3PO_4) react to produce calcium phosphate ($Ca_3(PO_4)_2$) and water. A. Write the balanced equation for this reaction. B. How many grams of calcium hydroxide ($Ca(OH)_2$) do we need to measure out to make 3.75g of calcium phosphate ($Ca_3(PO_4)_2$)? C. We made a mistake in our measurements. Instead of using the amount of calcium hydroxide you calculated in Part B, we used 2.37g of $Ca(OH)_2$ and 2.69g of H_3PO_4. How many grams of $Ca_3(PO_4)_2$ should be made under these conditions? D. Which compound is the limiting reagent for the reaction described in Part C? E. What would be the percent yield if we only collected 2.98g of calcium phosphate when we did the reaction described in Part C in a laboratory?

11. Contextualization of the test item:

a. Adding meaning by embedding contextual cues that help stage the test item as a problem.

b. Making connections between parts of the questions to establish flow.

c. Adding more steps to scaffold all parts of the problem.

An example below shows that to stage this problem better, the factory setting was added with a task assigned that is part of solving this problem "at your job."

Original	**Modified**
The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following: $N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$ If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?	You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below: $N_2(g) + 3H_2(g) \rightarrow 2NH_3(g)$ You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Data Collection

The process of data collection begins after the assessment items have been modified. Data is primarily collected in the form of student test responses and/or student interviews. A think-aloud protocol is typically employed during the interview process. A think aloud protocol is an important tool to get a glimpse of students' sense-making skills, obstacles they perceive to understanding or answering a problem, and their overall comprehension in real time. Think aloud protocols "provide a direct view of a reader's mental activity, a kind of window into those processes which are usually hidden" (Block, 1986, p. 464). Students are asked to read the question aloud and then talk through what they are thinking as they try to solve the question. In the Siegel et al. (2014) study, students talked about and wrote down notes as they solved the problems. During the think aloud process, researcher(s) can ask follow-up questions to students' responses.

At the end of the think-aloud protocol, post-assessment interviews are typically conducted with each participant. The purpose of post-assessment interviews is to allow students to reflect upon the quality of the item they had completed. This is a time when the students can discuss what they found most challenging or most helpful in an item, and/or to further clarify their understanding of an item. The main goal of including post assessment interviews in the Siegel et al. (2014) study was to determine the students' perceptions of the accessibility of the test items and how students would alleviate any perceived obstacles present in the test items.

Data Analysis

Data analysis in both of the studies that adhered to the guidelines of the EFCA has focused on comparing student responses to original and modified assessment items. Students' understanding of original and modified items is typically evaluated using rubrics and/or interview data. Rubrics and interview data are analyzed once a set of criteria is established for each assessment item. The criteria are generally focused on students' understanding of particular features of each item, for example, flow of language, vocabulary, framing, organization, etc.

Additional criteria regarding students' experienced difficulty in an item have also been implemented; these can include the following: (1) whether or not student's interpretation matches the intended meaning of the item; (2) if the student found it easy to interpret the question; and (3) if parts of the question posed potential barriers to students' understanding.

Emerging themes are typically established based on patterns in the coded data. Trustworthiness is established by incorporating multiple researchers, multiple data sources, and/or by consulting with language specialists. Previous studies informed by the EFCA have also employed member-checking during the post-assessment portion of the interview to establish trustworthiness of the data and of the findings (Siegel, 2007; Siegel et al., 2014).

Limitations of the Equity Framework

A major limitation of the EFCA is that it is newer in the field of educational assessment and still evolving. Because of this, a formal critique of the EFCA as a theoretical framework has not yet been published. However, in examining the framework and attempting to apply it to design the current study, I identified some potential limitations of using the framework to inform the design of research studies. In this section, I explain these potential limitations. I will discuss how these limitations have been accounted for in my study in the Methods chapter (Chapter 4).

The foundational principles of the EFCA were shaped by sociocultural philosophies of learning and science education; however, the connections between this framework and theoretical foundations in sociocultural perspectives could be strengthened. Fully understanding and applying the EFCA requires a deeper look at its theoretical foundation; however, the lack of historical context about EFCA's theoretical connections makes it difficult to fully determine how it is related to more established theoretical frameworks such as constructivism and sociocultural theory.

Because the EFCA is a new framework, there are only two empirical studies that have utilized the EFCA as a guiding framework. The patterns of data collection and analysis described in previous sections have been established primarily based on the design of these two studies. Because there have been a limited number of studies that have applied the EFCA, it is challenging to understand a broader scope of data collection and analysis techniques that can be adopted under this framework.

While it is clear that an important part of applying the framework is knowing how to use the suggested item modifications, the EFCA lacks clear, operational definitions of each type of modification. Little guidance is provided in previous studies on how each type of modification can be consistently applied to revise original test items. Because of this limitation, item modifications have been operationally defined for the current study.

Justification for Using the Equity Framework in the Current Study

Because I found chemistry assessment items difficult to interpret as an ELL student, I am particularly interested in ways that chemistry assessment items can be modified to make them more accessible for other ELL students. The EFCA provides guidelines for assessment item modifications that have been effective for middle school science students. I am interested in examining whether these same modifications—or others—are useful for an undergraduate chemistry student population.

An important goal of this study is to understand how undergraduate students perceive typical and altered chemistry test questions. The EFCA provides the guidelines to make assessment items more equitable for a diverse student population without watering down the content (not decreasing the content difficulty) and still challenging students to think critically about key concepts. I believe that the EFCA matches the aims of this study because it preserves high content standards for refining assessments, while still focusing on multicultural aspects of evaluating students' knowledge in ways that reduce potential sources of bias in assessment items.

CHAPTER 4 METHODOLOGY

Research Design

The purpose of this study is to investigate how individual chemistry test items are interpreted by ELL and native English speaking (NES) students. This study uses the lens of the Equity Framework for Classroom Assessments (EFCA) to modify general chemistry assessment questions in ways that should make questions easier to interpret for all students. This task required a deeper examination of the construction of the test items and asking the student participants to identify features of the items that they considered confusing or helpful in their interpretations of the items. A distinct aspect of this research is that it looks directly to students' perceptions of the items' characteristics, and does not only rely on the test maker's intended meanings for the items. This study provided a unique opportunity to examine test items from the point of views of both ELL and NES students, thus potentially addressing the gap in the literature regarding the assessment of undergraduate ELL students in general chemistry.

The following research questions have been developed to match this study's purpose:

1. What are English language learners' (ELL) and native English speakers' (NES) perceptions of typical general chemistry exam questions as compared with chemistry exam questions that have been modified according to the equity framework for classroom assessments (EFCA)?
    a. Which features of the questions do ELL and NES students perceive to be helpful?
    b. Which features of the questions do ELL and NES students perceive to be challenging?
2. What modifications do ELL and NES students believe would make chemistry exam questions easier to comprehend?

Target Population

One important goal of this study was to understand how students who are still developing their fluency in the English language manage to interpret general chemistry test items. Another goal was to ensure that test items that have been modified according to the guidelines of the EFCA are accessible to all students. Accordingly, participants for this study included two types of students: (1) students who are native English speakers, and (2) those who are still in the process of learning English. To be eligible as an NES participant for this study, a student must have been born in the U.S. and be a native English speaker. To be eligible as an ELL participant for this study, students must have met the following criteria:

- Must have been born outside of the U.S. and/or other countries where English is the national language,
- Must have a first language (L1) other than English,
- Must be able to read, write, and understand basic conversational English,
- Must have been residing in the U.S. for 10 years or less.

These criteria for the selection of ELL participants are based on Cummins' (1980, 1981, 1984, 2000) work on language acquisition in English language learners. ELL participants who are not born in countries where English is the national language and whose first language is not English will not have a strong background in the English language. Based on Cummins' (1980, 1981, 1984, 2000) work on language acquisition, basic interpersonal communication skills begin to develop in the first two years of being in the English-speaking country for incoming ELL students; however, it could take up to 8-10 years for incoming ELL students to develop cognitive academic language proficiency. Because I was interested in students' interpretations of chemistry exam questions, tasks that require the use of academic language, I chose to solicit participants that have basic conversational skills in English, but who were still in the process of developing CALP according to Cummins' (1980, 1981, 1984, 2000) timeline. Because this group of students could be potentially more susceptible to struggling with interpreting exam questions, their perceptions were valuable to the current study.

Participant Demographics

A total of 20 participants were recruited for this study, 10 students were native English speakers, and 10 students were English language learner students. The age range of the participants was 18 to 23. In terms of gender, there were three males and seven female students in the NES participant pool. There was one male, and nine female students in the ELL participant pool. The racial and ethnic background of the NES participants are as follows: two Caucasian students, two Filipino (second generation) students, two Hispanic (second generation) students, one African American student, one Chinese (second generation) student, one Hawaiian student, and one Armenian (second generation) student. The ethnic and racial backgrounds of the ELL participants are as follows: two Filipino students, two Thai students, two Hispanic students (Colombian and Venezuelan), one Indian student, one Chinese student, one student from Guam, and one Russian student.

Academically, all participants reported that their grades were at a C or better in their current chemistry course. Participants were enrolled in general chemistry because it was a prerequisite course for their majors or degree program. In order to protect the identity of participants, pseudonyms have been assigned to each of the participants.

Data Collection

Think aloud protocols and post-assessment interviews were employed as the primary tools for data collection in another study informed by the EFCA (Siegel et al., 2014). This study employed retrospective protocols and post-activity interviews as the main methods of data collection instead. In retrospective protocols, participants are given time to read and interpret the information presented to them before they are asked to discuss their perceptions of the information. Retrospective protocols were utilized instead of think aloud protocol to allow for more processing time for students to read and interpret the item and to avoid potentially influencing the processing of the word problem. Retrospective protocols have been successfully used in similar studies that ask students to discuss their thought process of an assessment item in this manner (Noble et al., 2012).

Specifically, I asked students to read the assessments items presented to them. Then, they were asked to circle words or phrases that they found unclear in red ink and to circle any words or phrases they found helpful in guiding their understanding to be circled in blue ink. Then they were asked to discuss what the problem was asking, and how they would set up the item. At this point, they were reassured that they were not being graded and that I was more interested in their thoughts about the item than in correct answers. In the following section, I briefly describe the assessment items that were selected for this study, and then I discuss the student interview guide and methods of participant recruitment.

Assessment Items

The assessment items used in this study focused on the concepts of limiting reactant and percent yield, which are considered foundational topics for general chemistry. These topics focus on quantitative relationships between the compounds involved in a chemical reaction.

In the text that follows, I provide an example of a limiting reactant word problem and an example of a percent yield word problem, as well as their solutions. I also outline what students need to know and do in order to solve the problems.

Limiting Reactant Problem: What mass of water forms when solutions containing 3.27 g of hydrochloric acid and 7.62 g of calcium hydroxide are mixed?

Solution:

Students need to write the balanced equation for the reaction described in the problem.

$2HCl_{(aq)} + Ca(OH)_{2(aq)} \rightarrow CaCl_{2(aq)} + 2H_2O_{(l)}$

Students need to perform the following stoichiometric calculation to convert grams of HCl into moles of water.

$3.27\ g\ HCl\ x\ \frac{1\ mol\ HCl}{36.46\ g\ HCl}\ x\ \frac{2\ mol\ H_2O}{2\ mol\ HCl} = 0.0897\ mol$ H_2O = **Limiting Reactant**

Students need to perform the following stoichiometric calculation to convert grams of calcium hydroxide into moles of water.

$$7.62\ g\ Ca(OH)_2\ x\ \frac{1\ mol\ Ca(OH)_2}{74.092\ g}\ x\ \frac{2\ mol\ H_2O}{1\ mol\ Ca(OH)_2} = 0.206\ mol\ H_2O$$

Once the limiting reactant is identified (reactant yielding the smaller number of moles), students must convert the smaller number for moles of water to grams of water.

$$0.897\ mol\ H_2O\ x\ \frac{18.006\ g}{1\ mol\ H_2O} = 1.62\ g\ H_2O\ forms$$

Percent Yield Problem: Phosphorus reacts with bromine to form phosphorous tribromide. If 35.0 g of bromine are reacted with 27.9 g of phosphorous tribromide are formed, what is the percent yield?

Solution:

Students need to write the balanced equation for the reaction described in the problem.

$2\ P + 3\ Br_2 \rightarrow 2PBr_3$

Students need to perform the following stoichiometric calculation to convert grams of bromine into grams of phosphorous tribromide to obtain the theoretical yield for the reaction.

$$35.0\ g\ Br_2\ x\ \frac{1\ mol\ Br_2}{159.808\ g\ Br_2}\ x\ \frac{2\ moles\ PBr_3}{3\ moles\ Br_2}\ x\ \frac{270.686\ g\ PBr_3}{1\ mol\ PBr_3} = 39.5\ g\ PBr_3$$

Once grams of phosphorous tribromide is obtained, students then need to use the equation shown below (inputting the experimental amount given in the problem and the theoretical amount obtained above) to calculate percent yield.

$$\%\ yield = \frac{experimental\ amount}{theoretical\ amount}\ x\ 100$$

$$\%\ yield = \frac{27.9\ g\ PBr_3}{39.5\ g\ PBr_3} * 100 = 70.63\%$$

I chose to use limiting reactant and percent yield assessment items for the current study because word problems associated with this content typically contain technical language and require mathematical computations, both of which have shown to be challenging for students (Carter & Brickhouse, 1989). I applied the following criteria in selecting the specific limiting reactant and percent yield assessment items for the current study:

- Items must have mathematical and chemical concepts and phrases that students need to think through; in other words, the problems cannot be solved by simply plugging values into a formula.
- Items should challenge students to think about underlying chemistry concepts.
- Items must be revisable based on modifications of the EFCA.
- Items must realistically be potential assessment questions in general chemistry.

In consultation with other instructors of general chemistry and according to criteria I established, I chose seven items for the current study. Original items were selected from general chemistry textbooks and a general chemistry instructor's test bank. Each item was revised based on the modifications (see Appendix A) derived from the EFCA (see Appendix B for all items and Appendix C to see how each item was modified). Each item was also assigned a content difficult level of easy, medium, or hard. The content difficulty levels were assigned by the original sources of the items, which were textbooks, and general chemistry instructors. Of the seven original items selected and revised according to the EFCA, four total items were used during the interviews because of time limitations. Items 1, 2, 5, and 6 were selected because they included a range of modifications (see Appendix A for a complete list). They also varied in content difficulty levels (e.g., easy, medium, hard). Item 1 was rated at medium difficulty, item 2 was rated as hard, item 5 was rated at medium difficulty, and item 6 was rated as easy.

Interview Guide

The main goal of the student interview protocol was to understand students' perceptions of general chemistry assessment items. Accordingly, the interview guide (see Appendix D) was developed to allow students to express their thoughts and perceptions about the original items and the modified items. The interview consists of four main sections: (1) informed consent (see Appendix E for informed consent forms); (2) rapport building questions; (3) retrospective protocol activity questions; and (4) post-activity questions. The interviews were semi-structured in nature: students' responses were followed-up with additional questions as deemed appropriate and the interviews were conversational. Each interview lasted approximately 1 hour.

Rapport Building Questions. The goal of the questions in this section was to solicit information about students' backgrounds and to help the participants feel comfortable with the setting and the interviewer. I started by asking about the participants' country of origin, first language, and experiences when moving the U.S. (if ELL) in order to gain information about their language and cultural backgrounds. I also asked participants about their experiences in their chemistry course, such as their favorite topic(s) and how they felt about their course exams and quizzes in order to transition the flow of the conversation to general chemistry assessment and the retrospective protocol.

Main Activity. This was the main activity portion of the interview, in which data was collected that relates to both research questions. The participant was given a single general chemistry item on a sheet of paper, along with writing utensils. Whether the participant received an original version of an assessment item or a modified version of an assessment item was randomized. Participants were informed that I would like for them to read the question presented to them. After a few minutes, I asked them to make annotations on the paper: to circle parts of the item that they found challenging in red and to circle parts that were helpful in blue. At this point, they were asked to discuss each of the parts they circled. I also asked participants to talk through how they would set up the problem. I reminded the students that they were not being graded. I asked them to describe their problem solving process in terms of why they started where they did, what the problem is asking them to do, and whether or not the question was easy or hard to follow. Based on their responses, follow-up questions were asked during this activity such as "how do you think you would change the item to make it easier to understand?"

The order in which students received the questions was randomized to reduce the likelihood of bias that students may carry should they notice any patterns between the original and modified versions. When the original version was first presented to the participant, the modified version of the same question was then shown to the participant without letting them know which version of the question is which. The student was then asked to review and comment on the modified version as they did on the original version. Likewise, when a participant first received a modified version of an assessment item, they were then asked to review and comment in a similar way on the original version of the same question. After the participant commented on both versions of the same question, I placed both versions of the same item in front of them and asked them to compare the versions and comment on which one they would prefer to encounter on an actual exam. Lastly, participants were asked to suggest their own modifications of the items to make it easier for them to follow. This procedure was followed until the student had seen both version of all four assessment items.

Post-Activity Questions. After the main activity portion was completed, the participants were asked a few general questions about all the problems they completed in both versions. I asked the participants to choose (1) the most difficult item to understand and (2) the easiest item to understand, and to explain their choices. They were also asked to express their thoughts on modifications that would make their chemistry exam questions easier to understand in general. Finally, they were given an opportunity to ask questions and add any additional thoughts to their interviews.

Interview Procedure

Interviews were audio recorded. Participants' written notes about each problem were saved and collected for reference during data analysis. I also took notes during the main activity portion in order to track responses and develop potential follow-up questions during the interview. Participants were provided with a set of three pens: black (for notes to set up the question), red (for indicating unclear parts) and blue (for indicating useful parts). All materials were collected at the end of the interview.

Participant Recruitment

I used purposeful sampling to select undergraduate students for this study. In purposeful sampling, participants are sought that specify a pre-determined criteria based on the research question (Creswell, 2007). I solicited participation from students enrolled in the first and second general chemistry courses (General Chemistry 121 and 122) at the University of Nevada, Las Vegas (UNLV). ELL and NES students who were currently enrolled in these courses at the time of data collection and were 18 years of age or older were eligible to participate. I obtained the syllabi for the General Chemistry courses, CHEM 121 and CHEM 122, from the instructors to determine when the topics of limiting reactants and percent yield were covered in class. I recruited participants from CHEM 121 after the exams that covered these topics in their classes to ensure that they had had the opportunity to study these topics. Because students in CHEM 122 learned these topics in a previous course, I did not have limitations on when I could recruit these participants. I recruited the CHEM 122 students at a time when it was convenient for the instructor.

I attended course lectures to inform students about this project, with the instructor's permission. A pre-written script (see Appendix F) was read in the classes, informing students about the purpose and goal of this project. They were asked to participate in an in-person, face-to-face interview where they will be solve several limiting reagent and percent yield problems (see Appendix B). Interested students were asked to write their email addresses and phone numbers on index cards that I handed out in class. Students were contacted via email and/or phone to set up a convenient time and location for the interviews. The goal was to obtain 5-15 participants, which is in alignment with a previous, qualitative study that utilized the EFCA (Siegel et al., 2014). For the current study, I recruited 10 ELL students and 10 NES students total.

Data Analysis

The audio-recorded data from all participant interviews was transcribed verbatim. Once the interview transcripts were produced, I began the process of coding data. I used a grounded theory approach to analyze the interview transcripts and searched for emergent themes related to the features in the items. This process was iterative and continued until no new themes were identified. Participants' written notes were also used for cross checking new categories in order to maintain the authenticity of emerging categories.

Each general chemistry item (both versions) was coded separately as each item differed in the type of modifications applied. Each item was first coded individually based on the features that were identified by participants as being helpful. Next, each item was coded individually based on the features participants identified as being challenging. I noticed that these features—whether identified as helpful or challenging—were similar and only differed negatively or positively. Because of this, I grouped the helpful and challenging coded features and searched for similarities. These coded features were combined into larger categories. The following categories emerged:

- Description of the problem
- Bullet points
- Chemical formulas
- Word choices (simple or complex wording)
- Syntax
- Length of the item
- Graph
- Illustration

- Phrasing
- Steps to solve the problem (guidance provided or lack of support)
- Sentence structure
- Type of the question
- Background information: relevant or irrelevant
- Equation
- Visual appearance
- Support (or lack of support) to solve the problem

Next, I grouped categories that were related to similar types of features. The following list shows how categories were grouped:

- Visual appearance and length
- Wording, syntax, phrasing, and sentence structure
- Diagrams, graphs, chemical formulas, equations and illustrations
- Contextual support and guided support for the content

Each of these groups corresponds to a major theme that characterizes the data: (1) macrostructure, (2) readability, (3) representation, and (4) scaffolding. These four themes are based in research that characterizes features of written text.

Macrostructure is a term adopted from literature on reading comprehension (Lacroix, 1999; Lo et al., 2016). Macrostructure is an important aspect of reading comprehension (Lo et al., 2016). Research on the cognitive processes of reading a text suggests that the way textual materials are presented to readers allows the readers to form coherent mental representations of the text. Features such as headings, spacing and organization influence how readers perceive the textual information (Lacroix, 1999). The theme of *macrostructure* was used to describe the features that contributed to the overall appearance and textual structure of the item.

Readability is a commonly used term to describe how written words are understood by readers. Word choices, phrasing, syntax and sentence structures can contribute to how well text is communicated and understood. Thus, these types of features are commonly assessed to measure the overall readability of written text (Pontus et al., 2017). Therefore, the *readability* theme includes features that contributed to how well the participants were able to comprehend the item.

Visual representations are widely discussed in literature as a significant feature that supports students' learning in science (Evagorou et al., 2015). Prior studies show that embedding graphs, images, graphic organizers, diagrams, photographs, models, symbols, etc. helps both ELL and NES students understand assessment items (Siegel, 2014). Thus, the *representation* theme was used to describe these types of features.

Scaffolds are features that provide content support in assessment items in the form of contextual signals and guided inquiry. Scaffolds played a significant role in decreasing the performance gap between ELL and NES students in middle school life science assessments (Siegel, 2007; Siegel et al., 2014). These features help students make connections between parts of the item or provide additional information that influences how students understand the context of the item. Thus, the *scaffolding* theme includes features related to content support, guidance, and contextualization in items.

Each of the helpful and challenging categories described above was assigned to a corresponding theme. A summary of the themes, as well as examples of helpful and challenging features included in each theme, is shown in Table 1. A detailed explanation of each of these themes with operational definitions is provided in the results and discussion chapter (Chapter 5).

Table 1. Overarching themes that describe students' perceptions of helpful and challenging features.

Theme	Categories	Examples of Helpful Features	Examples of Challenging Features
Macrostructure	▪ Superficial structures ▪ Visual appearance	▪ Spacing within the text ▪ Text broken down into smaller parts	▪ One block of written text ▪ Lack of separation within text
Readability	▪ Sentence structure ▪ Syntax ▪ Wording ▪ Voice	▪ Direct sentences ▪ Simple vocabulary ▪ Active voice	▪ Indirect sentences ▪ Complex vocabulary ▪ Passive voice
Representation	▪ Graphic organizers ▪ Visual elements	▪ Graphs ▪ Illustrations ▪ Chemical formulas ▪ Chemical equations	▪ Incomplete representations ▪ Unclear information within representations
Scaffolding	▪ Content support ▪ Guidance ▪ Contextualization	▪ Phrases that point to key information ▪ One step guides to the next ▪ Contextual cues that help understand the problem	▪ Lack of signals to identify key information ▪ No clear delineation of steps ▪ Lack of contextual cues to understand the problem

It is important to ensure that the data analysis process is reliable among multiple coders; this is has been referred to as inter-rater reliability. This was done by taking a small number of completed interview transcripts and asking a colleague to read through the transcripts and code the transcripts according to the categories and themes I had developed. If potential differences were found between my analysis and her analysis, we discussed these and we came to an agreement about any changes to the description of categories. This ensured that my own interpretations were consistent with the data and useful to others.

CHAPTER 5 RESULTS AND DISCUSSION

Test Item Features: Chapter Overview

My objectives for this project were to identify undergraduate students' perceptions of original and revised chemistry assessment items and ask students to suggest additional modifications that could improve each item. Specifically, Research Question 1 asked: what are ELL and NES students' perceptions of typical general chemistry exam questions as compared with chemistry exam questions that have been modified according to the equity framework for assessments? Research Question 2 asked: what (additional) modifications do ELL and NES students believe would make chemistry exam questions easier to comprehend? The assessment items selected for this study were focused on the topics of limiting reagent and percent yield.

The findings from both research questions are presented together in this chapter. This decision was made because of the way students responded during the semi-structured interviews. In general, while students were able to identify beneficial and challenging features of assessment items, they were less able to identify ways in which items could be further improved. The few students who did provide suggestions for assessment item improvement did so in reference to challenging features of the item. For example, a participant stated, "I think it's confusing when different units are shown in the same long phrase," to which the interviewer responded, "I see…how would you change this part to make it easier to follow?" The participant responded by saying, "I would break down the phrase into multiple sentences and separate those units." Thus, participants' suggestions for item improvement were directly tied to the features they identified as challenging in an item.

Based on my analysis of student responses, four major themes emerged that represent the features of the assessment items that students considered helpful and/or challenging. The four major themes that emerged from the data are: (1) macrostructure, (2) scaffolding, (3) readability, and (4) representation. The operational definitions and descriptions of each theme are provided below. Item-specific examples of each modification discussed in the themes can be found in Appendices B and C.

Macrostructure

Macrostructure is an important emergent theme in this study that describes the superficial structures of the assessment items. These features were characterized by how the item visually appeared on paper and the affective responses the item elicited. Specifically, macrostructure encompasses features that contribute positively or negatively to the written text's visual and spatial presentation, and organizational format. For example, an assessment item (not used in this study) with one long block of text is shown below. Based on literature findings, most students would respond negatively to this block presentation of an assessment item.

> When ethane (C_2H_6) reacts with chlorine (Cl_2), the main product is C_2H_5Cl, but other products containing Cl, such as $C_2H_4Cl_2$, are also obtained in small quantities. The formation of these other products reduces the yield of C_2H_5Cl. Calculate the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2, assuming that C_2H_6 and Cl_2 react only to form C_2H_5Cl and HCl. Calculate the percent yield of C_2H_5Cl if the reaction produces 206 g C_2H_5Cl.

On the other hand, research indicates that students would be likely to respond to the assessment item more positively if the text were broken up into smaller parts, as shown below.

When ethane (C_2H_6) reacts with chlorine (Cl_2), the reaction produces C_2H_5Cl as the main product. Other products containing chlorine are also produced in small amounts. These are called minor products. Minor products decrease the amount of C_2H_5Cl (main product) that is made in this reaction.

A. Write the balanced equation for the reaction between ethane and chlorine that produces C_2H_5Cl and HCl as products.

B. Using the balanced equation you wrote for Part A, find the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2?

C. If you conducted the reaction described in Part B in the laboratory and only collected 206g of C_2H_5Cl, what is the percent yield of C_2H_5Cl?

Macrostructure is acknowledged in literature to play a significant role in enhancing reading comprehensibility and supporting the communicative purposes of the written text (Carrell, 1982; Kilray, 2015). For the purposes of the current study, macrostructure describes the following features identified by participants: bulleted or lettered lists or a physical separation of the background information from the question. These features impacted the perceived accessibility of the item.

Readability

Readability describes the features of the assessment item that positively or negatively contributed to how well the participants comprehended the item as intended. Features such as sentence structure (direct or indirect), vocabulary (simple or complex), syntax (i.e., the arrangement of words in the sentence), and tense (passive or active) were characterized within this theme. Prior research underscores the significance of these types of grammatical structures on the overall readability of written text and, thus, on students' performance on standardized exams (Amos, 2009).

For the purposes of this study, the following operational definitions are used:

- Direct sentences are those in which the object receives the action of the sentence (e.g., where can I buy lab coats?), where the object is the lab coat and the action is "buy." Indirect sentences also have the object and an action; however, these sentences include additional information that is not essential to convey the meaning (e.g., *could you tell me* where I can buy lab coats?)
- Simple vocabulary includes words that are less formal and used in everyday talk (e.g., use), whereas complex vocabulary includes words that more formal and/or technical in nature (e.g., utilized).
- "Syntax" describes sentence construction. A sentence with simple syntax contains a single clause (a sentence containing a subject and predicate). For example, I don't like dogs. Sentences with more complex syntax contain multiple clauses that are put together with commas and/or conjunctions (e.g., I don't like dogs, and my sister doesn't like cats because they make her sneeze).

Representation

Representation describes visual tools embedded in the item to organize, consolidate and/or illustrate information. This theme includes both the positive and negative features of the representation(s) provided in assessment items. Literature strongly suggests that visual representations—such as pictures, imagery, illustrations, and graphic organizers—assist with the understanding of text by reducing the load on working memory and augmenting the comprehensibility of the written text (Boonen et al., 2014; Siegel et al., 2014).

Scaffolding

The scaffolding theme includes features of the assessment items that offer—or fail to offer—content support and guidance to solve the problem. In previous similar studies, scaffolding is broadly defined as supportive structures that help students comprehend, visualize, organize thinking, and elicit responses (Siegel et al., 2014). For the purposes of the current study, scaffolds include the following features: (1) content support, (2) guidance, and (3) contextualization.

Content support is an important scaffold because it helps students identify the most essential pieces of information in an assessment item. Content support encompasses features that help students identify specific, content-based information to solve for limiting reactant and percent yield word problems such as the quantities of reactants, or components of the chemical reaction, and/or distinguishing actual and theoretical yield. For example, a percent yield assessment item included the following phrase: “You found that 35.0 g of acetic acid is actually produced. What is the percent yield of acetic acid?” The first phrase provides content support by helping students recognize that they are given the actual yield and will need to calculate the theoretical yield to find the percent yield.

The scaffolding theme also includes guidance in the form of contextual cues and signals embedded in the item that help students think about the information available and connect that information to potential next steps to solve the problem. This process is described as “guided inquiry-based learning,” which leads students to collect key information, ask questions, and make connections to useful knowledge (Akkuzu & Uyulgan, 2017). For example, in the item shown below, the lettered parts first instruct students to write a balanced equation, then to use that equation to find the theoretical yield.

When ethane (C_2H_6) reacts with chlorine (Cl_2), the reaction produces C_2H_5Cl as the main product. Other products containing chlorine are also produced in small amounts. These are called minor products. Minor products decrease the amount of C_2H_5Cl (main product) that is made in this reaction.

A. Write the balanced equation for the reaction between ethane and chlorine that produces C_2H_5Cl and HCl as products.

B. Using the balanced equation you wrote for Part A, find the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2?

C. If you conducted the reaction described in Part B in the laboratory and only collected 206g of C_2H_5Cl, what is the percent yield of C_2H_5Cl?

Contextualization is also an important scaffold. Context helps students make connections between the content and the environment in which the content can be relevantly applied. Contextualization is generally provided in an item's background information, which is stated before the question portion. Context is provided to help students situate new ideas and comprehend the concepts embedded in the item. For example, in the item below, the factory that manufactures ammonia gas using the Haber process is the context provided to situate this problem.

You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:

$$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$

You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

In similar studies that employed scaffolding in assessment items, the aforementioned features (content support, guided inquiry, and contextualization) helped students focus on the underlying goal in the question and attend to key aspects of the question. These features also increased the students' comprehension of the items (Siegel et al., 2014).

A summary of each of these themes along with examples of each theme's helpful and challenging features is presented below in Table 1, which was originally presented in Chapter 4.

Table 1. Overarching themes that describe students' perceptions of helpful and challenging features.

Theme	Categories	Examples of Helpful Features	Example of Challenging Features
Macrostructure	▪ Superficial structures ▪ Visual appearance	▪ Spacing within the text ▪ Text broken down into smaller parts	▪ One block of written text ▪ Lack of separation within text
Readability	▪ Sentence structure ▪ Syntax ▪ Wording ▪ Voice	▪ Direct sentences ▪ Simple vocabulary ▪ Active voice	▪ Indirect sentences ▪ Complex vocabulary ▪ Passive voice
Representation	▪ Graphic organizers ▪ Visual elements	▪ Graphs ▪ Illustrations ▪ Chemical formulas ▪ Chemical equations	▪ Incomplete representations ▪ Unclear information within representations
Scaffolding	▪ Content support ▪ Guidance ▪ Contextualization	▪ Phrases that point to key information ▪ One step guides to the next ▪ Contextual cues that help understand the problem	▪ Lack of signals to identify key information ▪ No clear delineation of steps ▪ Lack of contextual cues to understand the problem

Although I prepared seven modified items, initial interviews suggested that there was not sufficient time for students to respond to all six items. Therefore, I selected four items to include during interviews. The four total assessment items used in this study: items 1, 2, 5, and 6.

In the sections that follow, I present the findings of the current study, as related to student responses to questions 1, 2, 5, and 6, in that order. I begin by presenting students' responses to the original version of an item, followed by students' responses to the revised item. Figure 11 shows the order in which the findings are presented. Within individual subsections, themes that emerged from the data (see Table 1) are presented in order from the most commonly reported themes to the least commonly reported themes.

I. Original Test Item
    a. Helpful Features
        i. ELL students' perceptions of helpful features, organized by theme
        ii. NES students' perceptions of helpful features, organized by theme
        iii. Comparison of themes: Helpful features
    b. Challenging Features
        i. ELL students' perceptions of challenging features, organized by theme
        ii. NES students' perceptions of challenging features, organized by theme
        iii. Comparison of themes: Challenging features

II. Revised Test Item
    a. Helpful Features
        i. ELL students' perceptions of helpful features, organized by theme
        ii. NES students' perceptions of helpful features, organized by theme
        iii. Comparison of themes: Helpful features
    b. Challenging Features
        i. ELL students' perceptions of challenging features, organized by theme
        ii. NES students' perceptions of challenging features, organized by theme
        iii. Comparison of themes: Challenging features

Figure 11. An outline of the order in which findings are presented for each assessment item.

Results and Discussion

In the following sections, the item features referred to by students are shown in black text. The remaining portion of the item is shown in gray text.

Item 1: Original Version

The original version of this item exam item (shown below) assesses students' ability to calculate a theoretical yield for a reaction, and then to use that theoretical yield to calculate a percent yield.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

In order to solve the problem, students must:

1. Write a balanced equation for the production of methanol from carbon monoxide and hydrogen gas.
2. Then they must use stoichiometric relationships to calculate the theoretical yield.
3. Once the theoretical yield is found, that value needs to be used to solve for the percent yield for the reaction.

ELL students' perceptions of helpful features in item 1 (original). ELL students found this question to be difficult to understand and most students were unable to successfully set up the problem. Each participant was asked to discuss any features in the test item that he/she found to be helpful. These features are discussed in the order of the following emerging themes: (1) readability, and (2) representation.

Readability. The most common features perceived to be helpful by ELL students are features that pertain to the readability of this item. Specifically, ELL students found the direct sentence structure (shown below) to be easy to interpret.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

They understood that one goal of the item was to calculate the theoretical yield of methanol. ELL participants reported that the last phrase of this item also had direct and shorter sentences, which was easy to follow.

> **Anastasia:** So I like this… I noticed that the part where they talk about the task is broken up into much shorter sentences for all of the [background] information, so I think that part is good just because you have some processing time in between sentences.

ELL participants also liked the word “actually” in this item, as it prompted them to make a connection to actual yield, which is needed to calculate percent yield.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Representation. ELL participants also found the inclusion of the chemical formulas of the compounds involved in the chemical reaction for the formation of methanol to be helpful visual representations.

Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

NES students' perceptions of helpful features in item 1 (original). Whereas ELL students identified features related to the readability of the text and the symbolic representations included in the item, NES students focused on helpful features that contributed to the readability of the item and which provided scaffolding for solving the problem. These features will be discussed in the order of the following prevalent themes: (1) readability, and (2) scaffolding.

Readability. Similar to the responses of ELL participants, NES students identified helpful features regarding the readability of item 1. NES students found direct sentence structures in the question portion of item 1 to be helpful. Eric mentioned, "That's straightforward; calculate the theoretical yield of methanol. It's just more simple, condensed down and that's what I like about it."

Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Additionally, NES students noted that the word "manufactured" helped them understand that carbon monoxide and hydrogen are needed to make methanol. Matt stated, "Manufactured? I think it was an important word for the question. Without it, you really wouldn't necessarily know what combining carbon monoxide and hydrogen could really produce." Another word that Josh found helpful is 'reacted': "I like that they used 'reacted' here instead of 'mixed' because whenever I see 'mixed,' it throws up signs for that I need to add something…" Josh was referring to adding things mathematically.

Scaffolding. The next set of features that NES students perceived to be helpful were related to the background information provided in this item. The description of methyl alcohol and how it is formed was considered to supportive because it helped them understand the chemical reaction discussed in this item, which provided content support.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Comparison of helpful features in item 1 (original). Both ELL and NES students identified readability features, such as direct sentence structures in the questions of the item, to be helpful. Unlike NES students, ELL students pointed out that they particularly appreciated the fact that chemical formulas were given after the name of each compound. NES students, on the other hand, found the written description of the methanol reaction to be helpful in this item. This difference suggests that ELL students relied more on visual structures such as chemical formulas, which were less language-dependent compared to NES students, who were able to use language-dependent features of the item such as the description of the reaction. This finding is in alignment with previous literature that reports that representations are especially useful in reducing the cognitive load for students with limited English proficiencies (Boonen et al., 2014). A summary of the themes pertaining to this item are shown in the Figure 12.

Item 1 Original: Helpful Features

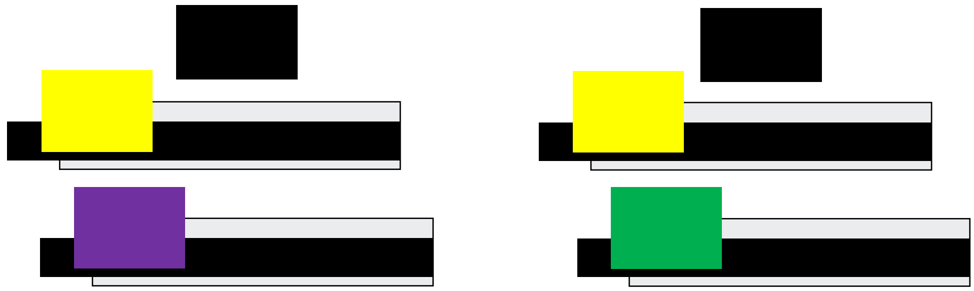


Figure 12. A comparison of the themes for item 1 (original) helpful features presented in the order of frequency.

ELL students' perceptions of challenging features in item 1 (original). ELL students indicated specific challenging features that could be improved in the item. The features that ELL students found to be challenging are related to the following themes: (1) scaffolding, (2) readability, and (3) macrostructure.

Scaffolding. ELL students highlighted concerns about the lack of contextual cues and guidance in the item, which left them uncertain about where to begin. This was especially the case with reading the first two sentences about the simplest alcohol and replacement for gasoline. ELL students struggled with understanding how this information was related to the main question, which asked about calculating the theoretical yield of methanol. They searched for a connection in the background information that could have help them solve the problem, which led to uncertainty and confusion.

Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Seojun: So I don't know if it's really needed to include this mess [pointing at the first two lines] that it is also called a methyl alcohol just because throughout the rest of the problem it's referred to as methanol. So it really doesn't matter what else you call it. Then I guess the same thing with this one with simplest alcohol it's not very relevant.

ELL students further indicated that they were confused by the amount of information that was embedded in the questions of the item. They explained that it is stressful to see multiple numerical values given in any one complex statement, especially when it is in the question portion of the item.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57×10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Lupe: Well, for me, you won't really know at first where to start. They just give you a bunch of numbers. It takes a while to figure out these numbers from this number and from this molecule, and then this one's from this. It's just hard to organize it at first. You actually have to work on it first before you really work on it or else you'll forget the details.

Readability. The next set of challenging features highlighted by ELL participants were those that made it more difficult for the students to comprehend the item itself. These features included word choices such as “potential” and “manufactured” which the students considered to be complex vocabulary. Students also felt that the item included long sentences containing complex syntax. For example, the question portion of the item below was written with hypotheticals words “suppose,” and “if,” to start the first part of the sentence followed by a direct question. This arrangement of words and phrases in sentences was perceived to be challenging to follow.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Rohan: I noticed that a lot of the word choices here are elevated and then in general seems like it has a longer sentence structure. ‘Potential’ and ‘manufactured’ are words that are sort of especially because some of this information is not super needed especially when it's talking about being simplest alcohol and like potential replacement for gasoline.

Rohan’s comment suggests that the use of “elevated” vocabulary in this item did not seem to be necessary, especially since these words are used to discuss parts of the item that were not considered to be important.

Macrostructure. ELL participants expressed verbally and in writing that they were overwhelmed as soon as they looked at this item because it appeared as one long block of text. They suggested separating the parts of the item to make it more visually accessible.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol?

NES students' perceptions of challenging features in item 1 (original). In general, the challenging features identified by NES participants were similar to those identified by ELL students. Despite finding challenging features in the item, most NES students were able to successful setup the problem. These features will be discussed in the following order of themes: (1) scaffolding, (2) macrostructure, and (3) readability.

Scaffolding. NES participants indicated that one of the most challenging parts of this item was that the background information was irrelevant to the questions that the item was asking. Many NES participants regarded the information about methyl alcohol and its use as a source of fuel to be unnecessary and advised that it be eliminated.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57×10^4g CH_3OH is actually produced, what is the percent yield of methanol?

The responses of NES participants also indicated that the fact that the item required two separate answers—the theoretical yield of methanol and the percent yield of methanol—was not clear. Many students assumed that they only needed to report the percent yield of methanol.

Alex: I'm going to go back and double-check…is it asking for two [answers]? Oh, calculate the theoretical yield. And then with the percent yield…

Interviewer: Do you think other students would have missed that?

Alex: Yeah. I definitely would've.

Macrostructure. Another challenging feature reported by NES students regarding the original item 1 is that appears to be one, long block of text, which contains a lot of critical information without visual cues, such as spacing between parts or a graphic organizer.

> **Nora:** Because with this one it looks like a paragraph of information. I feel that's how some of my chem questions are like, and this is when it starts to get hard for me because when I have so much information to sift through I don't know what I'm looking for.

> Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$, calculate the theoretical yield of methanol. If 3.57×10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Readability. NES participants identified word choices in the original item 1 that made this question difficult to follow. The use of ‘manufactured’ was questioned instead of simpler alternatives such as ‘made.’ Additionally, starting a sentence with the word ‘suppose’ was noted to be especially unsettling during an exam.

Nora explained why the use of “suppose” deterred her:

> **Nora:** Because people are really unsure and scared going through this and you hear, ‘Suppose,’ and couldn’t it just be stated like, ‘Yes, this is how much was put in.’ […] I’m already nervous and scared and your 'suppose' makes me nervous and scared for the rest of it. That’s just my own personal feeling, [especially] if my test anxiety is already shooting.

Comparison of themes: Challenging features in item 1 (original). Both ELL and NES students reported similar challenging features in original item 1. Both groups indicated that it was visually overpowering to come across a question like this during an exam, where the text appears to be in one long paragraph form. Both groups of students did not find the background information provided to be necessary and/or relevant. ELL students did not find the word "manufactured" helpful; however, some NES students mentioned that it was good word to use. ELL students noted that the use of complex words like "manufactured" was unnecessary, especially because when these words were not conveying essential information needed to solve the item. On the other hand, NES students considered the word "manufactured" to be important chemistry vocabulary that needed to be embedded in the item. This suggests that NES students see words such as "manufactured" as situational and understand that certain words should be used in certain contexts.

A summary of the comparison of themes is shown in the Figure 13.

Item 1 Original: Challenging Features

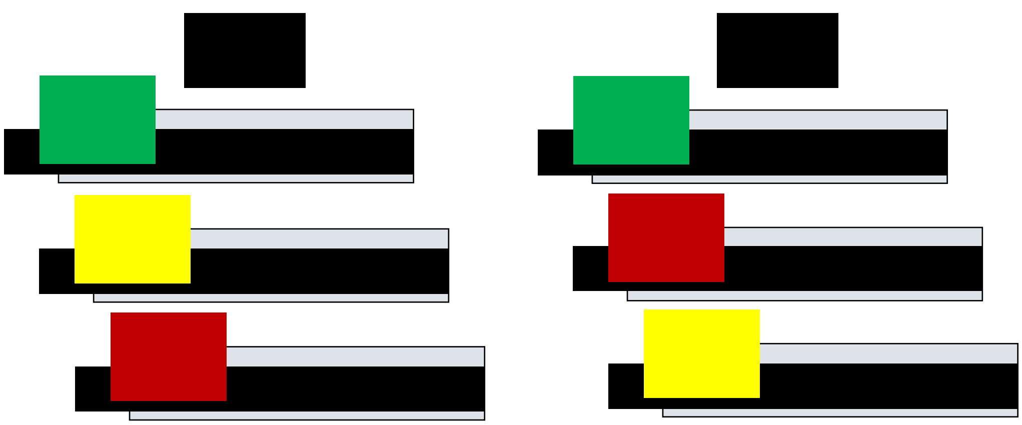


Figure 13. A comparison of the themes for item 1 (original) challenging features presented in the order of frequency.

Figure 13 indicates that NES students found the visual appearance of the item more challenging than the features that affected the readability of the item. ELL students, on the other hand, identified more readability-related issues as being challenging than macrostructure-related features. This implies that readability issues had a stronger effect on ELL students than they did on NES students. It is important to note that most ELL students were unable to set up this problem successfully while most NES students were able to set it up correctly. This may suggest that NES students were able to understand the complex words and sentence structure in the item enough to be able to set up the problem correctly. This finding is consistent with literature that reports that ELL students are more negatively impacted by language-dependent features than NES students on test items (Martiniello, 2009).

Item 1: Revised Version

The original version of Item 1 was revised according to the EFCA guidelines discussed in Chapter 3 and modifications discussed in Appendix A. The revised version includes modifications that linguistically simplified the background information by dividing complex sentences into simpler sentences and adding chemical formulas after the names of all compounds. The two questions were also separated by bulleted points. The revised version of item 1 is shown below.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

ELL students' perceptions of helpful features in item 1 (revised). ELL participants generally responded favorably to the modified features in the revised version and were able to successfully set up the problem. Data is presented below in the order of most prevalent emerging themes: (1) macrostructure, (2) scaffolding, (3) readability, and (4) representation.

Macrostructure. ELL participants found it especially useful that the revised item appeared to have distinguishable parts. For example, the space after the background information and the two bulleted questions were visually appealing and provided clarity. This was beneficial because many ELL students did not realize that the original item required two separate answers until they were able to look at the revised item, which made the two separate parts apparent. ELL students admitted that their answers to the original version of item 1 would most likely have been incorrect if it were on an exam. The helpful features in terms of macrostructure on the revised item 1 are shown below.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> [Space]
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

Scaffolding. ELL students perceived that the bulleted questions offered support and guidance to set up and solve the problem. They discussed that the separated questions provided content support because the first bulleted part prompted them to solve for the theoretical yield of methanol, which would be needed to later solve for the percent yield of methanol using the actual yield given. ELL students also acknowledged that the third sentence in the background information about how methanol is made was useful in enhancing their understanding of the chemical reaction taking place in the problem. One ELL student also pointed out that the phrase starting with "you found that…" resonated with her because "you" adds context to the item and "actually" signals that the value is for actual yield.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that 3.57×10^4g CH_3OH is actually produced. What is the percent yield of methanol?

Sheela: This one's [revised version item 1] much better, it's neater and you know which steps you need to take. It's much easier to understand because you can distinguish between which is the background, which is the main sentence. Then, the bullet points tell you which parts you need to do and then there's two steps to the problem.

Readability. Furthermore, ELL participants perceived that the direct sentence structures present in the questions of the item supported their comprehension. The bulleted parts of the item that start with "what" were perceived to be easy to read.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that 3.57×10^4g CH_3OH is actually produced. What is the percent yield of methanol?

Representation. Lastly, ELL students identified the chemical formulas next to the names of compounds methanol, carbon monoxide, and hydrogen gas as helpful visual features in the item. Chemical formulas provide a symbolic representation of each of the compounds discussed in the item. They provide a way to visualize a chemical reaction and make it easier for a student to write a balanced equation.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

NES students' perceptions of helpful features in item 1 (original). NES participants generally responded positively to the revised version of item 1. The features they found to be helpful have been thematically arranged and will be discussed in the following order: (1) macrostructure, (2) representation, and (3) readability.

Macrostructure. The most dominant helpful feature in the revised version of item 1 that the NES students mentioned was related to the visual organization of the question, especially the separation of the questions into two separate bullet points. NES students found this revised version to be structurally clearer and noted that it was easier to determine what they were asked to do in response to the item.

Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2). • What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$? • You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

Abby responded positively to this item when shown, “Obviously, it's cleaner looking … like … the two questions are broken up so you don't have to keep rereading it and looking for what was it asking. […] It's just easier to see.”

Representation. The next most frequently reported helpful features by NES participants regarding the revised item 1 were related to the theme “representations”: the presence of chemical formulas of the chemical compounds, methanol, carbon monoxide and hydrogen gas. Although the chemical formulas were provided in the original item, NES students mentioned that these representations were more clearly evident in the revised item because the chemical formulas appeared in parentheses next to the name of each compound as it was initially being discussed. Ana stated, “[chemical formulas] help me put this together, I have an image of what the reaction looks like.”

Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2). • What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$? • You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

Readability. NES students also indicated other helpful features in the revised version that augmented the overall readability of this item. Among these were the simple sentence structure used in the description of the reaction, which made it easier to read the background information provided. Compared to the original version of this item, the revised version modified syntax and simplified longer sentences into shorter sentences, which was perceived to be easier to follow.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that 3.57×10^4g CH_3OH is actually produced. What is the percent yield of methanol?

Scaffolding. NES students identified features in the revised item that provided guidance as being helpful. They specifically found that bullet points led them to first find the theoretical yield, and then to use the provided actual yield to solve for the percent yield. Emery mentioned, "This one was pretty easy because it's like writing the book. Each question leads to the other. You need each question to finish this whole series of questions." Ana similarly stated, "I like how you need the answer for each question to move on to the next." Jessica also mentioned, "It walks you and it guides you down a little more of what you're doing."

Comparison of themes: Helpful features in item 1 (revised). A comparison of themes between ELL and NES students for the revised item 1 suggest that both groups responded favorably to the overall presentation (macrostructure) of the item 1 (revised). However, ELL participants more frequently indicated that they benefitted from the structured bulleted parts, which gave them the guidance and a map of where to start and how to solve the problem than did the NES students. In terms of readability-related features, the groups of participants identified different helpful features: NES students found the simple and shorter sentences and syntax in the background information portion to be useful; whereas ELL students identified the direct sentences in the questions that start with “what” to be useful. Both groups reported that the chemical formulas were clearer to see and that the simple sentence structure made it easier to read the provided background information.

These findings suggest that both groups of students found visual organization an important feature in this item. The separation of the background from the bulleted parts seemed to help students see that this item consisted of a two-part question and enhanced the item’s accessibility. This point is in alignment with prior research that suggests that structuring written text spatially adds to visual clarity and also plays a role in reducing cognitive load (Lacroix, 1999).

Another important finding is that while ELL participants relied on scaffolding features when attempting to set up and solve the problem, NES participants reported using more language-dependent features, such as sentence structures and wording, to set up the problem. This suggests that students with lower English language proficiencies may particularly benefit from scaffolding within items, as reported in similar studies (Siegel et al., 2014).

A summary of the comparison of themes is shown in Figure 14 below.

Item 1 Revised: Helpful Features

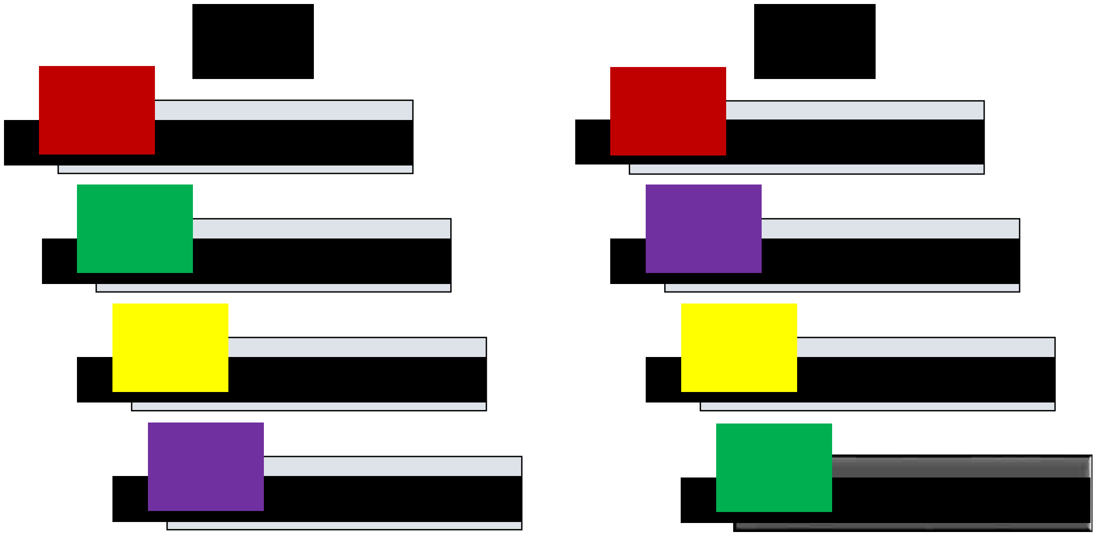


Figure 14. A comparison of the themes for item 1 (revised) helpful features presented in the order of frequency.

ELL students' perceptions of challenging features in item 1 (revised). ELL participants largely favored features of the revised version of item 1. Their responses indicating challenging features in the revised version of item 1 were limited to one aspect: scaffolding.

Scaffolding. The ELL participants identified that the background information about simplest alcohol and gasoline provided in the item was unnecessary, irrelevant and/or not helpful in solving the problem.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

NES students' perceptions of challenging features in item 1 (revised). Although NES participants also favored many features in the revised version of item 1, their responses indicated several challenging features, related to the themes of scaffolding and readability, which could be improved.

Scaffolding. NES participants identified that the background information provided in the item as unnecessary, irrelevant and/or not helpful in solving the problem. Abby stated, "I would remove the fact that methanol can replace gas just because I really don't think it helps with calculating your answer." Likewise, Ana mentioned, "Yes, take this part out – it's fun to know these little fun facts, but I mean in an exam (laughs), cool, but I don't need to know this." Ana's comment suggests that, during a timed exam, additional information not related to solving the problem is not helpful.

> Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).
>
> - What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
> - You found that $3.57x10^4$g CH_3OH is actually produced. What is the percent yield of methanol?

Readability. NES participants also indicated that the revised item contained confusing words that made it difficult to follow. The words 'mixing' and 'mix' were considered problematic because some students perceived 'mix' to suggest that the numerical values for carbon monoxide and hydrogen gas should be mathematically added together since these compounds are chemically reacting. Jessica expressed, "Whenever anything says mix, I immediately think of adding but I feel that's not what I'm supposed to do." NES students suggested eliminating "mix" and using a more precise word such as "reacted."

Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2).

- What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?
- You found that 3.57×10^4g CH_3OH is actually produced. What is the percent yield of methanol?

Comparison of themes: Challenging features in item 1 (revised). After reviewing revised and original items, both groups of participants reported that they would choose the revised version of item 1 on an actual exam. Both ELL and NES students did not find the background information provided in the revised item 1 to be useful, especially on an exam. The background information was included in the item to add context to the problem, which is recommended to be a useful feature in previous studies conducted with middle school students (Siegel, 2007); however, NES and ELL undergraduate students' responses suggest that any information that is not directly tied to setting up and solving the problem is not perceived to be useful.

Another interesting finding is that NES students identified readability-related features in terms of word choices that they found challenging. NES students indicated that the words "mixing" and "mix" were not as clear as "react," which is used in the original version. Many NES participants commented that "mix" can mislead students to think that the values of the reactants should be mathematically added and suggested that "react" is better. This is an interesting point because the word "react" is considered to be more technical than "mix" and it was intentionally replaced by "mix" in the revised version for the purposes of linguistic simplification (Siegel, 2007). This finding suggests that NES students are more experienced with gauging the technical accuracy of words compared to ELL students. Because of this, certain connotations of words should also be considered when changing vocabulary as it can influence the way students form interpretations.

A summary of these themes is shown in Figure 15.

Item 1 Revised: Challenging Features

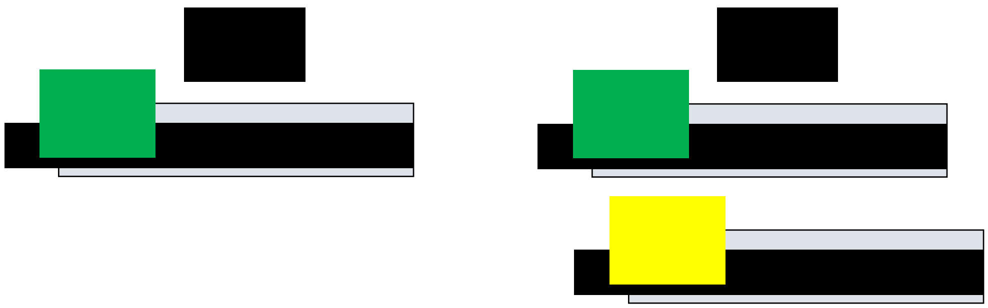


Figure 15. A comparison of the themes for item 1 (revised) challenging features presented in the order of frequency.

Item 2: Original Version

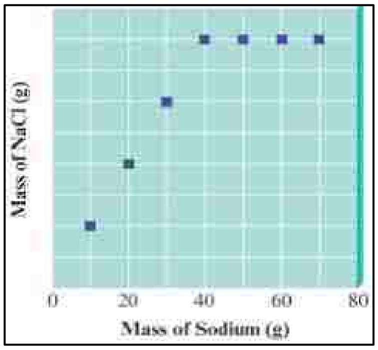
Item 2 (original) required students to interpret a graph to determine the limiting reactant in a reaction, using stoichiometric calculations (shown below). The following steps are required to solve this item:

1. To successfully solve this item, the student must recognize–without being prompted–that she/he needs to first write a balanced chemical equation based on the description provided about Na being added to Cl_2 to produce NaCl.
2. Next, part a asks the student to explain the shape of the graph, which requires her/him to read the graph provided and understand that the production of NaCl plateaus once 40 g of Na is added. At this point, they may notice that there are no numerical values provided on the y-axis which correlate with the amount of NaCl produced.

3. Part b expects the student to solve for the mass of NaCl made when 20.0 g of Na is added by performing stoichiometric calculations based on the balanced chemical equation.
4. Part c asks students calculate the mass of Cl_2 in each container. From the initial description provided about the seven containers, they know that there will be equal amount of Cl_2; however, to determine an exact numerical value of Cl_2, they must use a mass of Na and solve for the mass of Cl_2 using stoichiometric calculations.
5. Part d requires a similar setup of the problem to part b. The students needed to solve for the mass of NaCl produced when 50.0 g of Na is added by performing stoichiometric calculations based on the balanced chemical equation.
6. Finally, students are asked to identify the limiting reactant (Na or Cl_2) and calculate its mass in the previous steps, b and d. This part then requires that the masses of each compounds be stoichiometrically used to find a mass of the product, NaCl. The reactant yielding a lower amount of product is the limiting reactant.

The item 2 (original) was deliberately selected in this study because it has a higher level of content and conceptual difficulty as it was labeled as a "challenge question" in the original source of this item. This item requires higher problem solving skills than other limiting reactant items and the ability to make conceptual connections to identify the steps necessary to solve it. Students' responses to this item were helpful because they provided insights on the types of linguistic features students may rely on and/or find unsupportive in solving challenging limiting reactant word problems.

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.

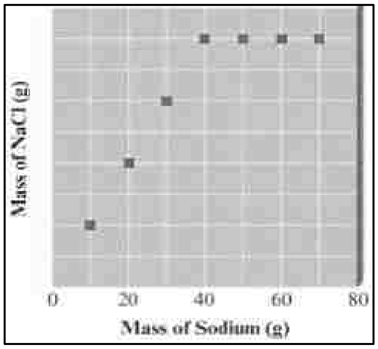


a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

ELL perceptions of the helpful features in item 2 (original). ELL students generally found this item to be difficult to solve and were unable to successfully set up the problem; however, they reported helpful features represented by the following themes: (1) scaffolding, (2) readability, (3) macrostructure, and (4) representation.

Scaffolding. ELL students indicated that they found the background information about the setup of the reaction to be helpful and relevant in understanding the overall context of the item. For example, the information about how the reaction of chlorine gas and sodium is set up tells the student to how to write a chemical equation.

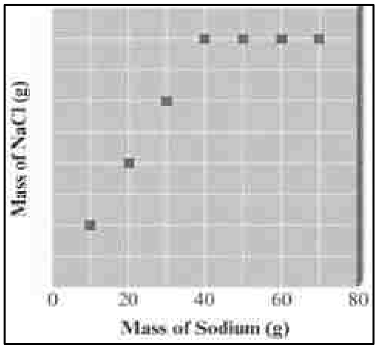
You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Readability. Additionally, ELL students reported that they appreciated having direct sentences in the lettered points of the item after the graph because these points were easy to read and comprehend. Ina commented, "…at least I know what they want…if they want a mass, they say, 'calculate the mass of'…"

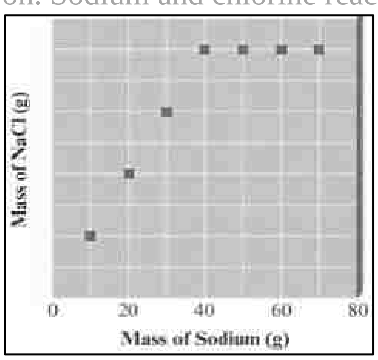
You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Macrostructure. ELL students reported that seeing that the question was organized in lettered points separated from the background made it appear more approachable. Elara mentioned, “I just think it’s not as scary looking. Plus, I kind of jumped to just looking straight at the graph, rather than going back and re-reading.”

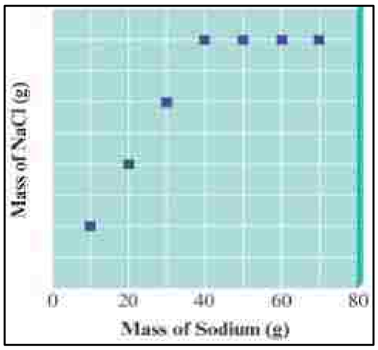
You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Representation. As mentioned by Elara above, ELL students specified that the graph was a positive feature of the question because they were able to see that as sodium was added to the reaction, it changed the amount of sodium chloride produced up to a certain point. Lupe stated, "I thought it was a really good visual representation of what was going on."

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



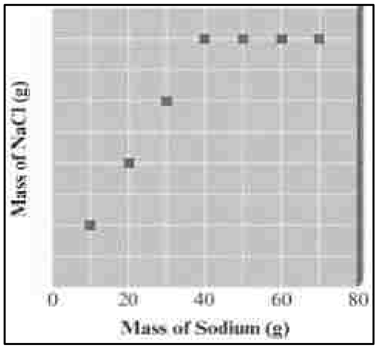
a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

NES perceptions of the helpful features in item 2 (original). Similar to ELL participants, NES participants found the original version of item 2 to be difficult to solve, and only a few students were able to set up this problem successfully. The key helpful features identified by NES students are represented by the following themes: (1) macrostructure, (2), readability, (3) scaffolding, and (4) representation.

Macrostructure. The more prevalent features identified by NES participants were related to the overall appearance of this question. Students noticed that the item was structured in a way that appeared to be shorter in length and that parts were separated into bulleted points. Abby was pleased to see that amount of reading required on this item minimal: "I liked [item 2 original] because you don't have to read too much." This comment is interesting because it seems that the format of the item made it appear that the item did not require as many steps as it actually does to set up and solve the problems.

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.

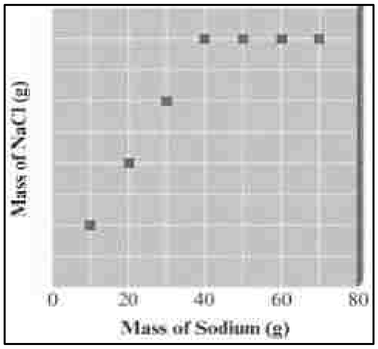
[Space]



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Readability. NES participants noted that this item was easy to read because it contained direct sentences in the bulleted parts. Emery mentioned: “I like how this one’s worded. ‘Calculate the mass…’ That’s straightforward.”

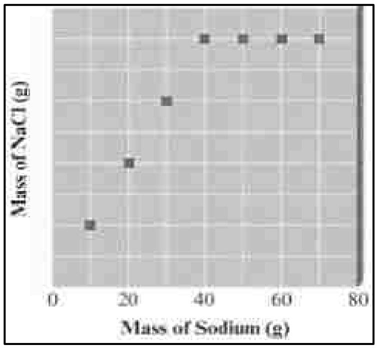
You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Scaffolding. The next set of helpful features discussed by NES students showed that they found the description of the reaction provided in the first part of the item to be helpful. Students mentioned that this information enabled them to better interpret the graph provided. Ana stated, “They are telling you that the first sample has this amount and the second sample has this amount. It just helps you understand what the graph is supposed to be.”

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

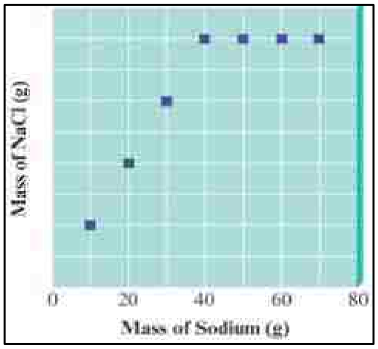
The description of the reaction also helped students understand that the same reaction is taking place in seven different containers.

Jasper explained the usefulness of this feature:

> **Jasper:** I think seven closed containers is pretty important because it tells you what you're looking at when you look at the graph. You're not just looking at somebody who just added to a substance to one container. It also tells you that the mass of chlorine gas does not change, that's important too…. Yet, sodium chloride amount peaks at some points and it's because the amount of chlorine gas didn't change. That's in the question so that is helpful.

Representation. NES students also indicated that having the graph available enabled them to understand the nature of the reaction described as they were able to track the amount of sodium added to the reaction and how it was related to the production of sodium chloride.

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Comparison of themes: Helpful features for item 2 (original). Both groups of participants found the original item 2 to be challenging to solve, which was expected. Although most students were unable to successfully set up and explain how to solve this problem in its entirety, they were able to point the features that helped them in the process of solving this item. ELL students reported that the description given in the beginning of the item was most helpful because the background information provided a description about how this reaction was set up over seven containers. NES students, on the other hand, found the most helpful feature to be the item's overall appearance. Both groups recognized that having direct sentences in the lettered parts of the item made it easier to read and know what the questions were asking. The graph was also considered to be a relatively helpful representation to convey the nature of the reaction.

An interesting finding here is that ELL students tended to identify scaffolding-related helpful features while NES students tended to identify macrostructure-related helpful features for item 2 (original). One possible explanation for this is that ELL students rely more on supportive structures that provide context to the problem they are being asked to solve (Siegel, 2007; Siegel et al., 2014). In the case of item 2 (original), the first three statements provided useful contextual information about how the reaction of chlorine gas and sodium was set up to form sodium chloride in seven containers. This information provided the foundation needed to interpret the graph. NES students may not have needed this description as much as ELL students since NES students indicated that they noticed the graph and the lettered parts of the item first. Literature on the relationship between language proficiency and academic performance suggests that ELL students need contextual connections in items that are more cognitively demanding, as this item was (Cummins, 1984).

Although both the ELL and NES participants determined that the background information provided in item 1 (original and revised) was not helpful because they could not use it to set up the problem. In item 2, however, the students found the background information helpful because they *could* use it to set up the problem. This suggests that there are differences in the type of contextualization employed in an item in terms of its usefulness for students. If the context supports problem solving, then it is considered helpful. If the context provides additional information that does not contribute to solving the problem, then it is not considered helpful.

A comparison of the themes that emerged from students' responses about helpful features in the original item 2 are shown in Figure 16.

Item 2 Original: Helpful Features

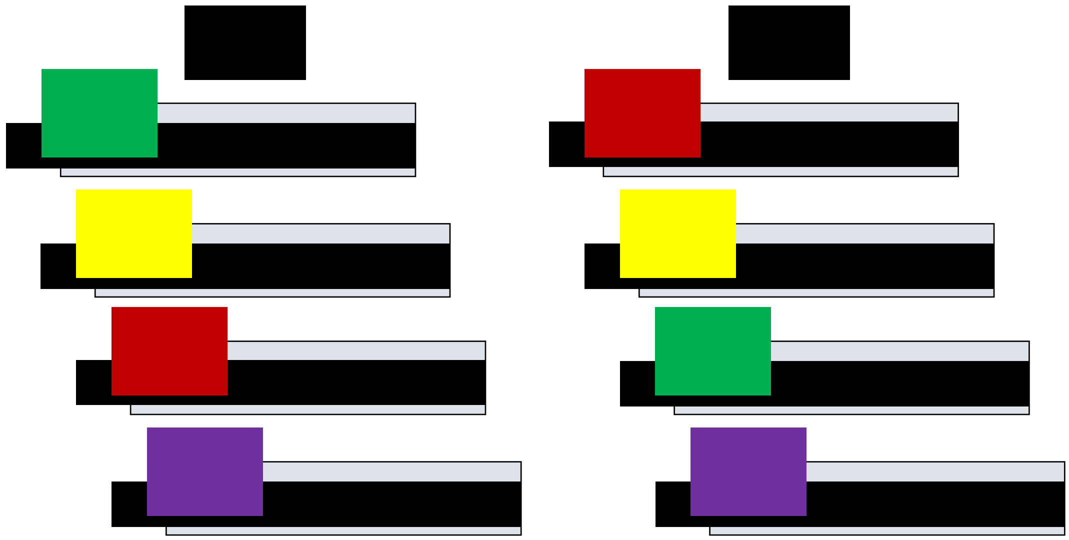


Figure 16. A comparison of the themes for item 2 (original) helpful features presented in the order of frequency.

ELL students' perceptions of challenging features in item 2 (original). ELL students identified challenging features that have been organized in the following themes: (1) representation, (2) scaffolding, and (3) readability.

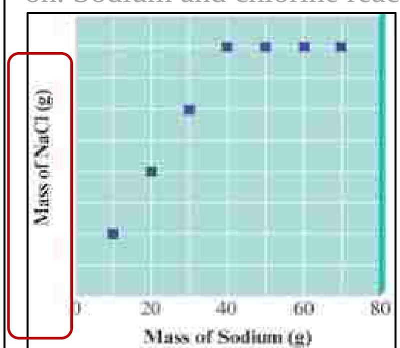
Representation. Although it was visually appealing to have a graphical representation included in the item that described the reaction of sodium and chlorine to form sodium chloride, not having a y-axis labeled was problematic part of calculations for students. ELLs were drawn to the graph; but, upon further inspection, they could not understand whether or not the graph was intentionally given without the y-axis.

Interviewer: Where do you think other students may have struggled in this question?

Naima: I think they would have a tough time trying to think about what the y-axis is, the numbers and the values for the y-axis, and that would just throw them off of just figuring out B, C, and D parts, and eventually towards E.

To successfully solve this problem, the students would have needed to determine the numerical values on the y-axis using stoichiometry to solve for the grams of NaCl formed in each container.

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

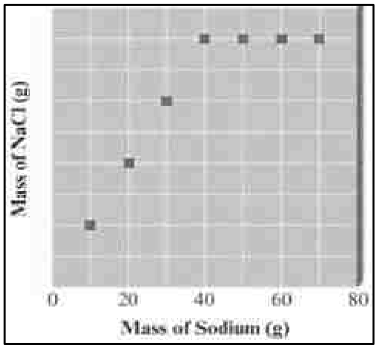
During the interviews, I observed that several ELL students drew their own visual diagram of the seven containers in the experimental setup. When I asked Lupe why she had drawn this figure, she responded, “It’s just easier for me.” Ultimately, when I asked the students how item 2 (original) could be improved, ELL students suggested that it would be helpful to include a visual representation of the background information about the reaction in seven containers.

Scaffolding. In addition to the confusing graph and the lack of a representation of the seven containers, ELL students also indicated that it was difficult to figure out how to begin solving the problem. After carefully reading all of it (and re-reading), ELL students struggled with finding an effective starting point for answering parts b, c, d, and e. Carlos stated, "…it took a couple of times to read because they asked for a lot of things so I just took a couple of times to read through it."

Rohan remarked that it was challenging to connect all the pieces of this item. He stated, **"**[item 2 original] is the hardest because it looks simple but the wording makes you think a lot, and people would get confused especially since you don't have any guidelines...confusing steps. There's no y-axis value and stuff."

Readability. The final set of features ELL participants found to be challenging were focused on vague wording and phrases in the item. Phrases such as "and so on" in the background portion of the item and "explain the shape" in part b were perceived to be vague. It was unclear whether or not the students should make the assumption that sodium continues to be added to in the same increments over seven containers or more. Additionally, the word "explain" in part was not precise enough to relay the type of information students should be conveying as part of their answers. For example, "explain the shape" could potentially be asking the student to discuss the shape that results from connecting the data points, the slope of the line, and/or the features of the graph such as axes, relationships, etc.

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Interviewer: Do you think students would be able to answer part a?

Seojun: Yes, especially if they're just—their first language is [English]. They probably just even write one word. But for me, I can't really put a shape to it right away. I just had to describe what it looks like to me.

NES students' perceptions of challenging features in item 2 (original). NES students also found the original item 2 to be challenging to solve even though they appreciated its structured presentation. During the interviews, NES students would often try to ask me follow-up questions to check their understanding and receive further guidance in order to clarify features they found challenging. The specific features NES students identified as challenging are arranged in the following themes: (1) scaffolding, (2) readability, (3) representation.

Scaffolding. The main challenge reported by NES participants was that the item was difficult to solve because it was difficult to determine which steps to take to solve the problems.

NES students felt that there was a lack of guidance and embedded cues that help them understand where to begin and provide reassurance along the way. For example, Alex mentioned, "… it doesn't remind you to balance the equation."

NES students also perceived that there are many implicit connections and steps they are expected to make, which is stressful during an exam. Nora expressed, "I feel like if you were to get [item 2 original] on the test, and you don't know what's going on, I feel like you would get anxiety…" Ana discussed how there was not enough guidance to understand the concept in the item. She mentioned, "This one [item 2 original], I could have solved the math. But I wouldn't have really understood the entire concept that they were going after […] but I definitely think that this one you don't learn from."

Readability. NES students also suggested that there were several words and phrases that could be clearer to communicate what the question is asking. For example, "remaining," "and so on," and "explain the shape" were perceived to be vague and elicited multiple meanings.

> **Abby** indicated: I personally just hate it when they're like "explain the shape" of the curve for me.
>
> **Interviewer**: Why is that?
>
> **Abby**: Well, I don't know, it's not -- Do you want it mathematically or…I don't know. I was like it's a positive slope, and then it levels off. Do you want a simple version? Or do you want me to go into details that there's a positive correlation?

Matt further echoed Abby's concern regarding part a.

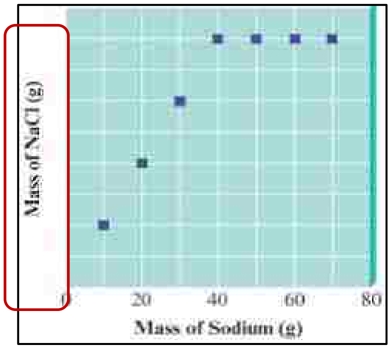
Matt: I feel like not every student would put down the same answer. I feel like if someone were to say shape like someone would say like an L, because sometimes when I see shape, I'm like "it's a circle." Maybe for this, it'd be better to describe the relationship or at least that's what I would put down.

Although "remaining" was not perceived as a complex word, students were unfamiliar with being asked about the "remaining reactant" since they were more used to being asked about the "limiting reactant," which caused some uncertainty. The phrase "so on" was considered vague because students are forced to assume that sodium is continually added in the same increments.

Representation. NES participants conveyed that while the graph should have been helpful, it appeared to be incomplete. They explained that they were unsure if the absence of values on the y-axis was deliberate or an error was made in the question itself, which caused confusion. They were unable to determine that they would need to perform stoichiometric computations in order to label the values on the y-axis.

Overall, NES students conveyed that additional clarification about the values on the y-axis would help them interpret the graph. When asked to make suggestions how to improve this item, most NES students reported that a visual representation about the setup of the reaction would have been a helpful tool in the question because picturing seven containers and how it relates to the graph was challenging. Alex explained his overall impressions about the original item 2 when he said, "…There's no visual aid for the containers […] and then, the y-axis…there are no units on there and no numbers."

You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.



a. Explain the shape of the graph.
b. Calculate the mass of NaCl formed when 20.0g of sodium is used.
c. Calculate the mass of Cl_2 in each container.
d. Calculate the mass of NaCl formed when 50.0g of sodium is used.
e. Identify the remaining reactant, and determine its mass for parts b and d above.

Comparison of themes: Challenging features in item 2 (original). The original item 2 was generally perceived to be challenging to solve for both ELL and NES participants. Although the students identified helpful features regarding the item's structured presentation, graph and bulleted points, both groups also perceived that it contained features that were problematic for understanding the item. ELL students' perceptions highlighted that they tried to infer guidance from the graph provided; however, this strategy was not effective because the graph lacked values on the y-axis. Because of the lack of content support and signals, ELL students could not make the connection that they needed to use the masses of Na to solve for the mass of NaCl in order to assign values to the y-axis. Both ELL and NES students indicated the presence of complex vocabulary deterred their ability to comprehend the question as intended.

A summary of the order of emerging themes between ELLs and NES students is presented below in Figure 17.

Original Item 2: Challenging Features

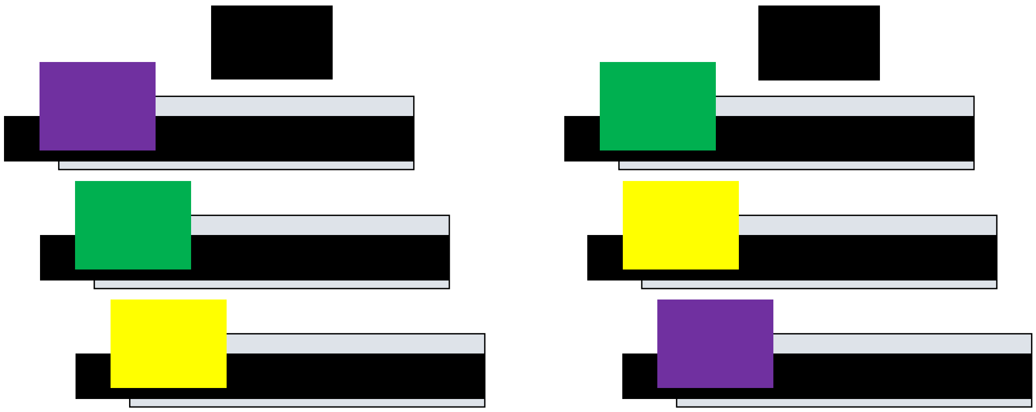


Figure 17. A comparison of the themes for item 2 (original) challenging features presented in the order of frequency.

Participants' responses imply that each group approached the item differently in trying to make sense of how to begin solving the problem. ELL students read the background description of the reaction first, which they found to be helpful; next, they looked to the graph to understand how adding sodium is related to the production of sodium chloride since the amount of chlorine added remained the same. Unfortunately, the graph led to uncertainty and confusion as ELL students could not figure out how or why the y-axis values were missing and that they needed to solve for the mass of sodium chloride using stoichiometry. ELL students looked at the lettered parts for content support and guidance on how to begin solving the problems, which was also lacking.

NES students, on the other hand, first approached the lettered parts of the item to begin solving the problem; however, they noticed that there was little guidance about how to begin setting up the calculations. NES students were then confused by complex words and phrases, followed by the missing information on the graph. This finding suggests that ELL and NES students look for different types of features to make sense of and solve difficult, multi-step word problems. Consistent with the literature (Siegel et al., 2014), the ELL students prioritized using the representation to form their interpretations of the problem, as this is a less language-dependent strategy for developing understanding of a problem.

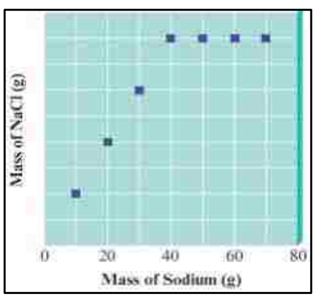
Item 2: Revised Version

The original item 2 was modified based on the guidelines of the EFCA discussed in Chapter 3 and modifications discussed in Appendix A. The revised version of item 2 was modified to include two visual representations of the setup of the reaction. The first showed seven reaction containers, and the second was a list of the masses of Na in each container. The lettered subparts (a,b,c,d,e) in the original item were rewritten and restructured to provide scaffolding by embedding guided inquiry and content support intended to help students make connections from one step to the next in solving the item. For example, part a in the revised version instructs the students to write a balanced equation, which was not explicitly mentioned in the original item. Part b in the revised item notifies the student that there are intentionally no numbers on the y-axis because they are expected to compute the mass of NaCl based on known values of Na. The next eight steps similarly offer content support and guide them to solve for Cl_2 and, eventually, identify the limiting reactant. Because of the length of this item, which spans two pages, the item has been divided into two parts as shown below: (1) the background information part, including the jars, list and the graph; and (2) the lettered parts.

There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows:

Jar 1: 10.0g $Na_{(s)}$
Jar 2: 20.0g $Na_{(s)}$
Jar 3: 30.0g $Na_{(s)}$
Jar 4: 40.0g $Na_{(s)}$
Jar 5: 50.0g $Na_{(s)}$
Jar 6: 60.0g $Na_{(s)}$
Jar 7: 70.0g $Na_{(s)}$

Sodium and chlorine react to make sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data.



Answer the following questions about this reaction and the data in the graph:

a. Write the balanced equation for the reaction between Cl_2 gas and sodium.

b. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in Part A, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2?

c. Based on the amount of NaCl that you calculated in Part B, label the y-axis of your graph.

d. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?

e. Use the balanced equation from Part A to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5.

f. How does the amount of Cl_2 you calculated in Part E compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?

g. What is the mass of the leftover reactant in jar 5?

h. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following:
    i. Why does the amount of NaCl produced increase at first when you add more sodium?
    ii. Why is the amount of NaCl produced constant when you add more than 40g of sodium?

ELL students' perceptions of helpful features for item 2 (revised). As the revised version of item 2 was presented to ELL participants, they were generally more confident in approaching it and solving each part of it. They also reported that they would be more likely to get the revised item 2 correct if it were on an exam. When asked to identify the features of the item they found to be helpful, ELL students reported the features related to the following themes: (1) scaffolding, (2) representation, and (3) readability.

Scaffolding. The most helpful feature reported about the revised item 2 was the content support and guidance it provided for solving the problem. Although ELL students found this item difficult to solve, ELL students were able to begin setting up the problem and successfully explain it. Many ELL students discussed that they were able to follow the logic of steps a through h as they were thinking through this item because these steps helped them connect the computational parts involving stoichiometry with the graph to the concept of limiting reactant. **Ina** stated, "It leads you, it guides you. Each question leads to another one."

> **Sheela**: Even though there are much more words, it's more comprehensive than the last one. The other one is just like, Oh, what is that? But you don't really know where to start with. This one it gives it to you like … for here, you tell me to write a balanced equation, it's what I did. You can solve that relationship, you see that you're supposed to use this relationship to figure out the rest…

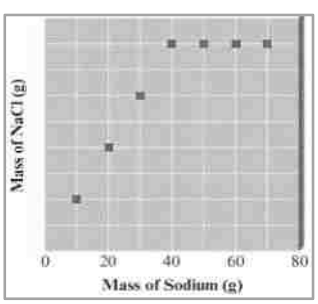
Representation. The next most commonly reported helpful features about the revised item 2 related to the provided visual representation of the seven jars, as well as the list of contents in each jar. Students discussed that it is often hard to remember the details given in the background of the item when trying to solve it, and being able to see an illustration of the jars helped them remember it without having to reread the background information. These representations can be seen in the first part of the item shown below.

Hector: Because you have a visualization of the parts of it. With the other one [item 2 original], I forgot that there's supposed to be seven [jars]. I totally forgot in the other one [item 2 original]. It just says so on and so on, I was like, Okay, I guess they just invented it.

There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows:

Jar 1: 10.0g $Na_{(s)}$
Jar 2: 20.0g $Na_{(s)}$
Jar 3: 30.0g $Na_{(s)}$
Jar 4: 40.0g $Na_{(s)}$
Jar 5: 50.0g $Na_{(s)}$
Jar 6: 60.0g $Na_{(s)}$
Jar 7: 70.0g $Na_{(s)}$

Sodium and chlorine react to make sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data.



Readability. Finally, ELL students discussed that although the revised version was longer in length than the original version, it was easy to read and comprehend. Because of the item's simple sentence structure and word choices, ELL students were able to more readily understand the tasks being presented to them. For examples, ELL students particularly appreciated reading the lettered parts that began with directives such as "write…" and "what is…" and "how…" These features have been shown in the latter portion of the item below.

Answer the following questions about this reaction and the data in the graph:

a. Write the balanced equation for the reaction between Cl_2 gas and sodium.

b. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in Part A, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2?

c. Based on the amount of NaCl that you calculated in Part B, label the y-axis of your graph.

d. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?

e. Use the balanced equation from Part A to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5.

f. How does the amount of Cl_2 you calculated in Part E compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?

g. What is the mass of the leftover reactant in jar 5?

h. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following:
   i. Why does the amount of NaCl produced increase at first when you add more sodium?
   ii. Why is the amount of NaCl produced constant when you add more than 40g of sodium?

NES students' perceptions of helpful features. NES students were more successful in their setup of the revised version of this item than they were in their setup of its original version.

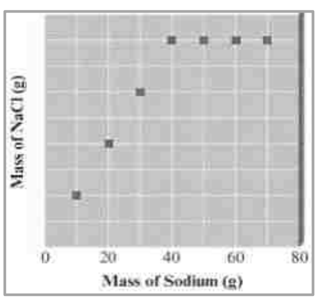
They also generally responded positively to the modifications that were added. The helpful features most commonly reported by NES students were related to the following themes: (1) representations, (2) scaffolding, and (3) readability.

Representation. The helpful features most commonly reported by NES participants were the visual representations of the jars and the list of the contents of each jar. The students discussed that having the visual representation of the jars was particularly helpful in understanding the setup of the problem. Emery stated, “I think that's [visual of jars] really helpful because you understand that each jar has the same amount of Cl_2 in it. It actually makes you understand.” These features have been shown below in the first part of the item.

There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows:

Jar 1: 10.0g $Na_{(s)}$
Jar 2: 20.0g $Na_{(s)}$
Jar 3: 30.0g $Na_{(s)}$
Jar 4: 40.0g $Na_{(s)}$
Jar 5: 50.0g $Na_{(s)}$
Jar 6: 60.0g $Na_{(s)}$
Jar 7: 70.0g $Na_{(s)}$

Sodium and chlorine react to make sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data.



Scaffolding. NES participants also identified helpful features in the revised version of item 2 that provided guidance for each step need to solve the problem. NES participants discussed that this question was particularly good at walking them through the logic of the problem and helped them understand why the values on the y-axis were left out. Nora commented, “This one is more specific because it guides you through the process. It starts off with ‘write your equation first’ [but] the other one [item 2 original] didn't mention any of that. That was on your own to figure out.”

Alex also added that the revised version was a good question for eliciting learning, which is important. However, he perceived that the original version of the item may be more appropriate for testing purposes because he felt that chemistry exam questions are not typically designed with this much content support and guidance. Alex believed this in part because he did not need as much guidance to successfully solve this item; he was able to solve it without relying on the added scaffolding.

> **Alex:** For learning something, I feel like this one 2 revised shows if you're proficient with it just because it's better for learning just because the few steps on 2 original are broken into almost double the steps in 2 revised. If a student's trying to learn it and wanted not a step-by-step but easing their way through it, 2 revised would be better, but 2 original is the better purely test question.

Readability. Finally, NES participants discussed that the revised version of item 2 was easier to comprehend because of its direct sentences and simpler words. For example, statements in the lettered parts such as "write the balanced equation," "what mass of NaCl," and "label the y-axis" are examples of what the students identified as helpful, direct sentences. These features are shown below in the first part of the item.

Answer the following questions about this reaction and the data in the graph:

a. Write the balanced equation for the reaction between Cl_2 gas and sodium.

b. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in Part A, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2?

c. Based on the amount of NaCl that you calculated in Part B, label the y-axis of your graph.

d. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?

e. Use the balanced equation from Part A to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5.

f. How does the amount of Cl_2 you calculated in Part E compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?

g. What is the mass of the leftover reactant in jar 5?

h. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following:
    i. Why does the amount of NaCl produced increase at first when you add more sodium?
    ii. Why is the amount of NaCl produced constant when you add more than 40g of sodium?

Comparison of themes: Helpful Features in item 2 (revised). Unlike the original item 2, ELL and NES students found the revised version of item 2 to be easier to read and solve. Although the revised version asked the students to solve for the same points as the original version, the revised version contains significantly more content support. ELL students found this type of support to be the particularly helpful in solving the problem. While NES students acknowledged that the scaffolding-related features provided guided support, they reported that visual representation of the jars to be more helpful than the content support.

The difference in the order of the emergent themes emphasizes the helpfulness of scaffolding particularly for ELL students on items with higher content difficulty. This finding is in alignment with previous research (Siegel, 2007; Siegel et al., 2014) that indicates that the use of scaffolding is especially beneficial for ELL students and effective at lowering the test score gaps between ELL and NES students on science assessment items. Some NES students found that adding a high level of content support and guidance would make this item more appropriate for learning instead of for testing purposes. This viewpoint suggests that students may have preconceived expectations of exam items versus practice or homework items. Students may not expect chemistry exam items to contain guided inquiry, especially in college-level introductory chemistry courses, where most instructors utilize multiple choice formatted questions focused on computational problem solving (Hartman & Lin, 2011). As such, students may expect more content support and guidance to appear on homework items than on exam items.

A summary of the comparison of themes is shown in Figure 18.

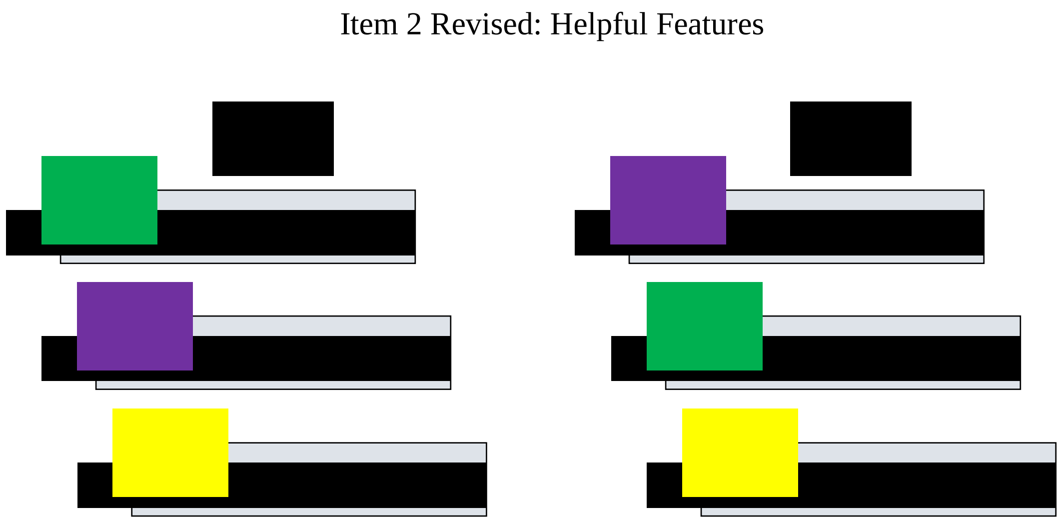


Figure 18. A comparison of the themes for item 2 (revised) helpful features presented in the order of frequency.

ELL students' perceptions of challenging features in item 2 (revised). ELL participants generally perceived the revised version of item 2 to be more accessible and easy to understand than the original version. When asked to identify challenging features within the revised version, they discussed how several features of the item could be improved to better suit their assessment needs. The major challenging features identified by ELL participants are related to the item's (1) macrostructure, (2) readability, and (3) representation.

Macrostructure. The main concern for ELL participants regarding the revised version of item 2 is that the item appeared to be lengthy at first glance. Compared to the original version of item 2, the revised version does have added steps and two additional visual representations (the illustration of the jars and the list of the contents of the jars), which required this item to be presented over two pages. ELL students discussed that when questions appear to be long and wordy, it is concerning because it requires more reading. Naima described her reaction, "Yeah, that's pretty overwhelming considering how the presentation to like all the questions are in one page but I think there are ways to make it shorter there's some stuff that I didn't think it was necessary.” Elara expressed a similar perception: “Okay, this one looks more overwhelming but easier to solve then the other one. I like the pictures here too.”

Readability. Furthermore, ELL participants identified features in the phrases and sentences in the item that they found difficult to interpret. They discussed that parts b and d had multiple big ideas in each step, which could have been separated or better organized to make simpler. As they were reading the parts, they paused and often looked up to ask for clarity by asking me if part b meant that they would need figure out all the numbers of the y-axis or just a few points. They felt that the instructions could have been clearer here. For part d, many students asked me about clarifying the second sentence. They found it awkward that it was stating that we did not know the mass of Cl_2, but we did know there was enough. These features are shown below in the latter part of the item.

Answer the following questions about this reaction and the data in the graph:

a. Write the balanced equation for the reaction between Cl_2 gas and sodium.

b. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in Part A, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2?

c. Based on the amount of NaCl that you calculated in Part B, label the y-axis of your graph.

d. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?

e. Use the balanced equation from Part A to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5.

f. How does the amount of Cl_2 you calculated in Part E compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?

g. What is the mass of the leftover reactant in jar 5?

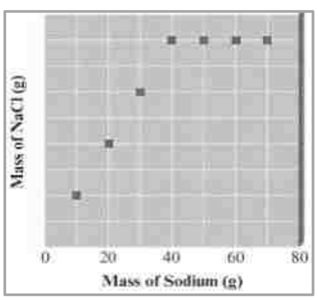
h. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following:
   i. Why does the amount of NaCl produced increase at first when you add more sodium?
   ii. Why is the amount of NaCl produced constant when you add more than 40g of sodium?

Representation. ELL participants appreciated having a visual representation of the setup of the problem. However, several participants suggested that it was unnecessary to have the visualization of the jars in addition to the list of the contents of the jars. They recommended that one of the representations (either the jars or the list) should be removed and it would reduce the length of this item. Some students recommended keeping the jars and others chose the list. The representations are shown below in the first part of the item.

There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows:

Jar 1: 10.0g $Na_{(s)}$
Jar 2: 20.0g $Na_{(s)}$
Jar 3: 30.0g $Na_{(s)}$
Jar 4: 40.0g $Na_{(s)}$
Jar 5: 50.0g $Na_{(s)}$
Jar 6: 60.0g $Na_{(s)}$
Jar 7: 70.0g $Na_{(s)}$

Sodium and chlorine react to make sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data.



NES participants' perceptions of challenging features in item 2 (revised). NES students' responses to the revised version of item 2 were similar to those of the ELL students. NES students echoed the concerns of ELL students in that although the revised item 2 is more accessible, it appears to be structurally lengthy and offered suggestions to improve the item. The challenging features they identified were related to (1) macrostructure, (2) readability, and (3) representation.

Macrostructure. NES participants perceived the most challenging feature in the revised version item 2 to be its length. The item appeared to be very lengthy and although the actual steps involved were easy to follow, the length made students nervous. Eric described his reaction to the item, "I think just having it so long and wordy, even bulleted A through J, it's a lot. I think it could overwhelm a student." Alex also expressed his concerns, "I just think it's too many words to read. When I read exam question, I just want to know straight what the information is, or what it's asking, and then the givens."

Readability. Additionally, NES students discussed how certain parts in the question seemed wordy and could be shortened to improve readability. For example, it was suggested that part c be connected as an extension to part b since part b was already referring to labelling the y-axis. Part d was considered to be redundant because the background information about the same amount of Cl_2 in all jars was reiterated here. These features are shown below in the latter part of the item.

Answer the following questions about this reaction and the data in the graph:

a. Write the balanced equation for the reaction between Cl_2 gas and sodium.

b. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in Part A, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2?

c. Based on the amount of NaCl that you calculated in Part B, label the y-axis of your graph.

d. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?

e. Use the balanced equation from Part A to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5.

f. How does the amount of Cl_2 you calculated in Part E compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?

g. What is the mass of the leftover reactant in jar 5?

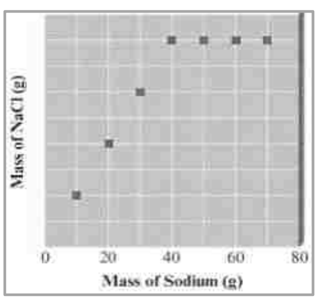
h. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following:
   i. Why does the amount of NaCl produced increase at first when you add more sodium?
   ii. Why is the amount of NaCl produced constant when you add more than 40g of sodium?

Representation. NES students also indicated that having two representations (the jars and the list) reflecting the same setup is unnecessary. By circling the visual representation of the jars and the list of the contents of the jars, they indicated that only one of these two is sufficient to convey the information that the amount of Na is changing, and the amount of Cl_2 is staying the same. Some students recommended keeping the jars and others chose the list.

There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows:

Jar 1: 10.0g $Na_{(s)}$
Jar 2: 20.0g $Na_{(s)}$
Jar 3: 30.0g $Na_{(s)}$
Jar 4: 40.0g $Na_{(s)}$
Jar 5: 50.0g $Na_{(s)}$
Jar 6: 60.0g $Na_{(s)}$
Jar 7: 70.0g $Na_{(s)}$

Sodium and chlorine react to make sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data.



Comparison of themes: Challenging features for item 2 (revised). Both ELL and NES students responded similarly in regard to the challenging features of revised item 2. Participants were able to follow the guided steps to setup the item. Most students from both groups also indicated that they would prefer to see the revised item version of item 2 on an exam. Students perceived that while the revised item was more accessible and easier to follow than the original version, item 2 revised can be further improved. The main challenge that both groups of students identified is the overall length of the revised item. The macrostructure of the item seemed to overwhelm students initially. Additionally, both groups perceived that having a visual representation of the jars as well as a list of the content of each jar was unnecessary and that only one representation would have been sufficient. It should also be noted that both groups of participants suggested having one of the two representations (the seven jars or the list) would be sufficient for their understanding. However, the preference for one over the other were evenly split among participants, which suggests that the type of helpful representation is a subjective choice for individual students regardless of language proficiency.

Another noted concern was the readability of parts b, c, and d. ELL students discussed that it was challenging to understand parts b and d because these parts contained too much information and, therefore, they were hard to process. They suggested separating the big ideas. By comparison, NES students stated that parts b, c, d were too wordy, redundant, and should be made more concise so that the information could be conveyed more efficiently. This suggests that although both groups found similar parts of the item hard to follow, NES students were still able to interpret the information and identify language-based changes that could be made to improve readability. ELL students, on the other hand, were unable to come up with concrete language-based suggestions to improve these item parts, which could be attributed to their developing English language proficiencies. As shown in prior research, this finding affirms that ELL students could be differentially and disproportionately affected by linguistically challenging factors on science assessments (Hartman & Lin, 2011; Turkan & Liu, 2012).

A comparison of the themes discussed has been shown in Figure 19.

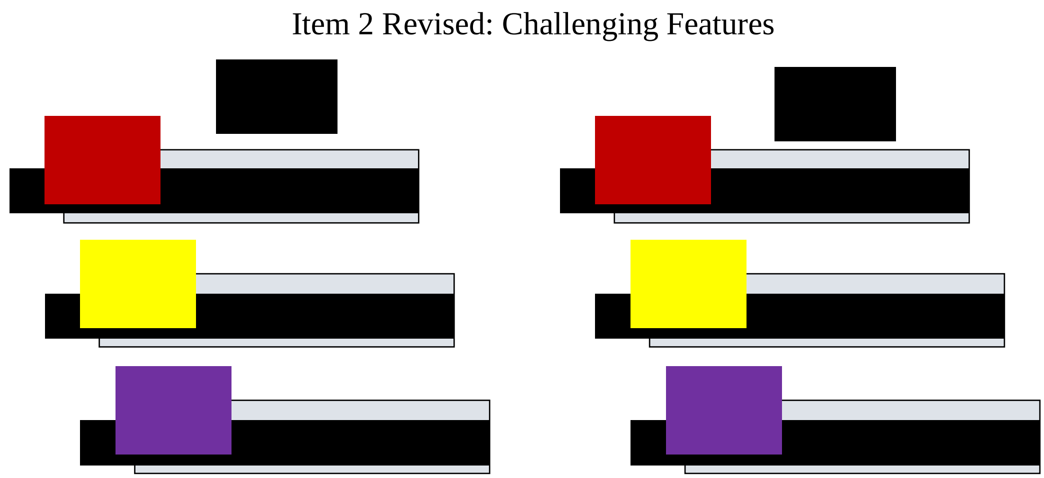


Figure 19. A comparison of the themes for item 2 (revised) challenging features presented in the order of frequency.

The students' responses to this item indicate that although there are advantages of adding content support and guidance to an item, there are other important factors to be considered such as the item's length and the amount of reading time it requires, especially on an exam. This was not something that has been examined in previous studies that implemented the EFCA with middle school students (Siegel, 2007). However, this factor is significant for undergraduate students at the university setting because most exams are time-limited.

Item 5: Original Version

The original version of this assessment item (shown below) asks the students to compute a mass of nitrogen based on the reaction of the Haber process. In order to solve this problem, students are required to:

1. Stoichiometrically calculate a mass of N_2 gas based on the given amount of 700g of NH_3
2. Account for the 70% yield by adjusting the mass of N_2

The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:

$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$

If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

ELL students' perceptions of helpful features for item 5 (original). ELL students generally responded favorably the original item 5 in that they seemed to understand it; however, they were unable to successfully set up the problem. Although students recognized that solving the item was more mathematically complicated than other items they had seen in the study, the item itself was easy to follow. To that end, ELL participants identified helpful features related to the following themes: (1) macrostructure, (2) readability, and (3) representation.

Macrostructure. The most commonly identified helpful features for this item were related to its short length and overall presentation. ELL participants discussed during the interviews that it was comforting to see that this item appeared to be short in length. ELL students said that short problems require less reading time, which is important to them on timed exams. They also indicated that the visual separation of the background information from the question in the item was helpful.

Readability. Additionally, ELL students identified helpful features for this item that were related to its simple sentence structure. ELL students found it especially helpful that the first sentence was not wordy and remained straightforward because it was directly phrased and contained simple syntax. They also reported that it was clear to follow what the question was asking them to do. Rohan's statement exemplifies this perception: "…for [item 5 original], the thing I like about it is that it's very forthcoming, it's good how they just go straight to the question and it stated directly what the problem is asking." Seojun similarly mentioned, "[This one is] somewhat simpler, it's just like the unnecessary stuff is basically taken out."

The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

Representation. ELL participants also circled and commented on the presence of the balanced equation provided in this item. This feature was considered to be a helpful representation of the Haber process as it shows that nitrogen gas reacts with hydrogen gas to produce ammonia. Students appreciated that the chemical equation was already balanced as indicated by the coefficients in front of the compounds. Another benefit of the provided chemical equation was that it showed the chemical symbols of the compounds, which was also considered useful by ELL students.

The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

NES students' perceptions of helpful features for item 5 (original). Similar to ELL students, NES students found this item to be succinct and easy to interpret; however, they had a difficult time setting up this problem successfully. They reported helpful features that were related to the themes of (1) readability and (2) representation.

Readability. NES participants appreciated that this item was readable and easy to comprehend. They indicated that words in this item such as "must produce," were simple and effective at pointing to the end goal of the item. They also identified the simple sentence structure of the first sentence to be helpful because it was easy to understand.

The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

Representation. As did the ELL students, NES participants noted that that the balanced equation was a useful feature. They discussed that the balanced equation provided in the item was a key useful feature because the equation shows the reaction as well as the chemical symbols of the compounds involved in the reaction.

The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

Josh: The equation is probably the most important part because even though the question tells you that the conversion of nitrogen and hydrogen can turn into ammonia, if they don't give you the formulas, it'd be hard to figure it out.

Comparison of themes: Helpful features for item 5 (original). The original version of item 5 was not considered to be an easy problem to solve computationally; however, both groups of participants identified key features that were helpful. ELL students primarily found the overall appearance to be helpful, while NES participants primarily found its simple vocabulary and sentence structure to be easily readable. The comparison of the emerging themes between ELL and NES participants is illustrated below in Figure 20.

Item 5 Original: Helpful Features

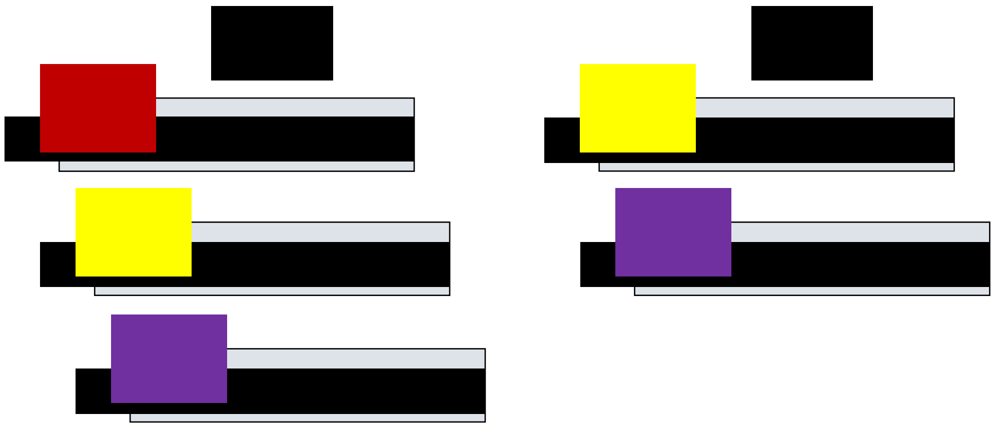


Figure 20. A comparison of the themes for item 5 (original) helpful features presented in the order of frequency.

These results suggest that ELL students were initially drawn to the item's organization and visual structure. ELL students realized that this item is short in length and required less reading and interpreting of the English language. NES students were more focused on the readability-related features, which were more language-dependent. NES students found the phrase "must produce" useful in directing them to the main goal of the item. By contrast, ELL students did not notice this phrase and only a few students identified the direct sentence structure of the question, which begins with "what mass of nitrogen…" to be useful in focusing their attention on the question. This suggests that NES students are more prone to recognizing specific words in items than ELL students, who tend to first focus on language-free structures.

ELL students' perceptions of challenging features in item 5 (original). The main difficulty they encountered was the inability to figure out whether the mass of N_2 should be decreased or increased to account for the 70% yield. In light of this, the challenging features they mainly perceived dealt with not having enough content support and guidance in the item. ELL perceptions of challenging features were related to the following themes: (1) scaffolding, and (2) readability.

Scaffolding. ELL participants discussed that the item lacked guidance on how to set up the problem and sounded vague. The background information provided was not perceived to be helpful and left students questioning how the term “high pressure” was related to the problem. Another concern expressed is that it was challenging to figure out where to begin the problem; some participants began with the 70% yield and others started with 700 g of NH_3. Lupe expressed her concerns, “This one's way more vague, and the way that they're asking you to find the mass of nitrogen is confusing.”

> The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
> $N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
> If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

Ina: I found really interesting where it describes the Haber process. I've never heard of this before but if people know or if the students know what this process is then would [they] know how to set up this problem? But at the same time, it’s kind of redundant because so I would say like instead of like giving this extra information, they would just might cut it out like the first sentence and then just include these two things like the chemical formula and reaction region and then the question directly beneath it.

Readability. ELL participants also perceived the last statement, where the main task is asked, to be confusing. Students suggested that this sentence be broken down into several simpler sentences instead of one, long complex sentence. Lupe’s suggestion reflects this notion, "Instead of making multiple sentences, they have a really long sentence that basically asks you to find the mass of nitrogen but it doesn't clearly point out which parts of the sentence are important."

> The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
> $N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
> If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

NES students' perceptions of challenging features for item 5 (original). NES students expressed significant concerns about the item containing background information that they did not find relevant. They also indicated features in the item related to readability, which they found to be confusing. Their perceptions are discussed below in the order of (1) scaffolding, and (2) readability.

Scaffolding. NES students indicated that the original version of the item included information that was not relevant to the task at hand. This was identified as a challenging feature because it made students wonder if that information was supposed to be somewhat essential to solving the problem. Additionally, students discussed how it was difficult to determine how to compute the 70% yield value once they found the mass of nitrogen.

> The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
> $N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
> If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

Ana expressed her concern with the phrase 'high pressure.'

Ana: There's nothing here that indicates that it's at high pressure. If there was maybe a symbol or something for high pressure, then that would be necessary so we could know what it's talking about. But it just looks like a simple reaction here.

Abby discussed that challenges of having the irrelevant background information.

Abby: At first, I thought it was going to be something like that was relevant to the question, but it’s not. It just doesn’t add anything, but I guess it’s a short phrase. It’s important to know what kind of reaction this is. But obviously having it on a test question is not more important, talking over lecture or something, but on a test question it just looks cluttered. It just gives you more to read and if you happen to think that you have to take into account high pressure while you’re doing this, then that would be important, because in physics if you talk about high pressure that you’d use a different equation what you have for interaction molecule order.

Jessica’s concerns were related to the order of information in the item.

Jessica: I guess for me personally, I like it when the question is stated at the beginning and then you have your givens versus having the question at the end. Because sometimes I’ll find myself running to finish, like on a test I want to get through it quickly. If I have the question here and we're told what to find first, then that's already in the back of my mind as I read the rest of the question. Just because it’s the way I process information.

Readability. NES students also identified parts of the item that were unclear. They indicated that the last sentence, which is the main question of the item, was a long, complicated sentence, which they found confusing. They discussed how a direct question would have been clearer in this case. Additionally, they suggested not placing information about two different compounds in the same, complex statement because it is confusing to understand which compound is being asked about. For example, the last sentence provides the mass of ammonia, but it is asking about the mass of nitrogen, which can be difficult to process.

> **Abby:** I kept repeating it over and over again because what was throwing me off was that at the end it was saying the percent yield of this reaction is 70%. My mind was thinking, What do I do with this?

> The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:
> $N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$
> If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?

Comparison of themes: Challenging features for item 5 (original). Both ELL and NES participants perceived this item to be difficult to solve. Both ELL and NES students perceived similar features to be challenging in the original version of item 5. The most commonly reported challenging feature was that description of the Haber process not relevant and useful for solving the problem. Many students mentioned that the information about high pressure deterred their understanding because they expected pressure to be involved in their calculations. Both groups also felt that the last sentence could have been simplified by being broken down into multiple sentences to make the question clearer and provide better content support. An illustration of a comparison of the emerging themes is shown in Figure 21.

Item 5 Original: Challenging Features

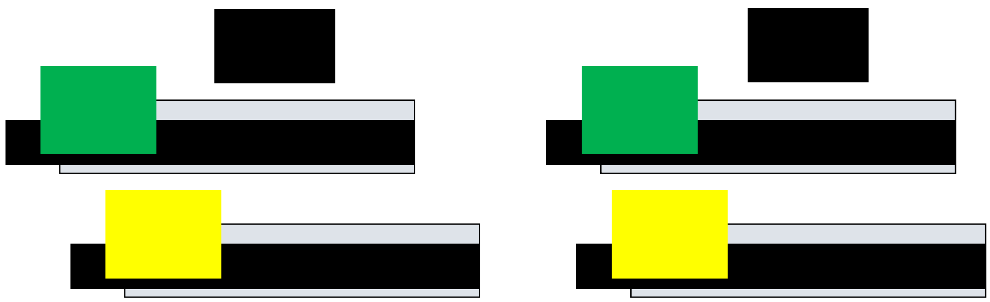


Figure 21. A comparison of the themes for item 5 (original) challenging features presented in the order of frequency.

An interesting finding here is that although both groups of students mentioned that this item was generally easy to follow and they understood what was being asked, most students were unable to correctly compute an answer for this item. Many participants were unable to first recognize that the 700 g of ammonia was an actual yield value, which speaks to chemistry content related issues that were most likely exacerbated by the challenging features identified by participants. These results suggest that along with the aforementioned challenging features, students' lack of chemistry knowledge and/or practice with solving this type of percent yield item may have contributed to their inability to solve the item successfully. This item required students to recognize that the 700g of ammonia is an actual yield, which needed to be divided by 70% to get the theoretical yield of ammonia to be produced. It is this value of ammonia that must be used to solve for the mass of nitrogen gas that should be used in the reaction.

Another important finding is that many participants believed that including background information that is not directly helping them solve the question should be eliminated on exam items. Participants' reasoning behind their perceptions indicates that their mindsets for exam questions is solely focused on the bottom line, which is adequately solving the problem within the time allotted. This type of mindset reflects that students carry preconceived expectations of exam items, which may be different if the item was presented as a homework item. This is an important topic that needs to be further investigated.

Item 5: Revised Version

The original version of item 5 was revised based on the guidelines of the EFCA as discussed in Chapter 3 and using the modifications shown in Appendix A. Specifically, contextualization was added to this item by including a storyline about the student working in a factory where the Haber process was used to make ammonia. The student in the storyline was tasked with finding the mass of nitrogen is needed to make 700 g of ammonia, with a percent yield of 70%. Additionally, the original version was revised by changing the sentence structure in the first and last sentences into direct sentences with simpler syntax. The presentation of the revised version was restructured by adding spaces between the background information, the balanced equation, and the question. The revised item 5 is shown below.

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

ELL students' perceptions of helpful features in item 5 (revised). When ELL students were asked to identify helpful features in the revised version of item 5, they generally responded positively to the modifications in the item. ELL students' perceptions of helpful features are discussed below in the order of the following themes: (1) scaffolding, (2) representation, (3) macrostructure, and (4) readability.

Scaffolding. ELL students mentioned that it was helpful to be told the objective of the question directly. They discussed that the part of the storyline where they were told that it was "your job" to make ammonia was particularly helpful as it functioned as a contextual signal that emphasized the question in the item. Additionally, ELL students also reported that this item provided more guidance that helped them follow the logic of the question.

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Hector discussed why he found these features helpful.

> **Hector**: Because they're telling you what you need to make. The whole story is that you are a factory worker, [chuckles] so, they give you what you need to make, which is an important part of the question in order to find the mass of nitrogen.

Representation. Similar to the original version of this item, the revised version also included a balanced chemical equation. This was an important helpful feature identified by ELL students because it showed the chemical symbols of each compound in the Haber process reaction and coefficients to indicate that the equation is balanced.

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Macrostructure. ELL students noticed that the revised version of item 5 included more spacing between the background, chemical equation and the question portions of the item. This separation of sections was considered to be helpful, especially if this item was given on an exam. Naima discussed this feature, “Putting space between parts or listing each part of the question or making just the question itself into multiple sentences really helps like in [this one].”

Readability. One of the challenging features discussed in the original item 5 was the length of the final statement. In the revised version, this statement was replaced with multiple shorter sentences that lead to the question. ELL students perceived this change in sentence structure to be helpful in understanding the question. Elara stated, “In [original] question 5, they just asked you one sentence, one really longer sentence with just the actual problem hidden in between, but in [revised] question 5, they split it into three sentences, and you knew what you needed to do.”

You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:

$$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$

You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

NES students' perceptions of helpful features in item 5 (revised). With the exception of macrostructure, NES students generally reported the same helpful features identified by ELL students. These features were related to the following themes: (1) scaffolding, (2) readability, and (3) representation.

Scaffolding. The most frequently reported helpful feature by NES participants related to the content support present in the revised item. The embedded contextual cues simplified and emphasized the question portion of the item. Specifically, NES students mentioned that the phrase "your job is…" provided a signal to focus on the main goal. Emery expressed, "I think the biggest part for me is when it says 'Your job.' Starting there, it solidified what I needed to do for this question." Jasper also stated, "It focuses my attention on what I need to do."

You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:

$$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$

You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Readability. NES students also indicated that the direct question and the short sentence structure in the question were helpful features. Compared to the original item, which contained one long statement in the end, this version divided the information over three shorter sentences with simpler syntax and added a separated direct question starting with "what." Nora mentioned, "I like the question, and the last sentence, 'what mass of nitrogen'…"

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Representation. NES students found the balanced chemical equation to be another helpful feature. Matt discussed that seeing the equation clarified the problem, "But after the equation is written, it's a lot more clear that what it's wanting you to do in terms of the steps that you're going to need."

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Comparison of themes: Helpful features in item 5 (revised). Both ELL and NES students identified similar helpful features in the revised item 5. The most commonly reported helpful feature was the content support provided in the revised item by including contextual cues and emphasizing to students that they are tasked with making 700g of ammonia. Both groups discussed that having direct and shorter sentence structures clarified the main point of the question. Additionally, both groups of students appreciated having the balanced chemical equation provided in the item. ELL students mentioned that the revised item was easier to approach because it had added spacing, a fact which was not identified by NES students.

A comparison of these themes has been shown in Figure 22.

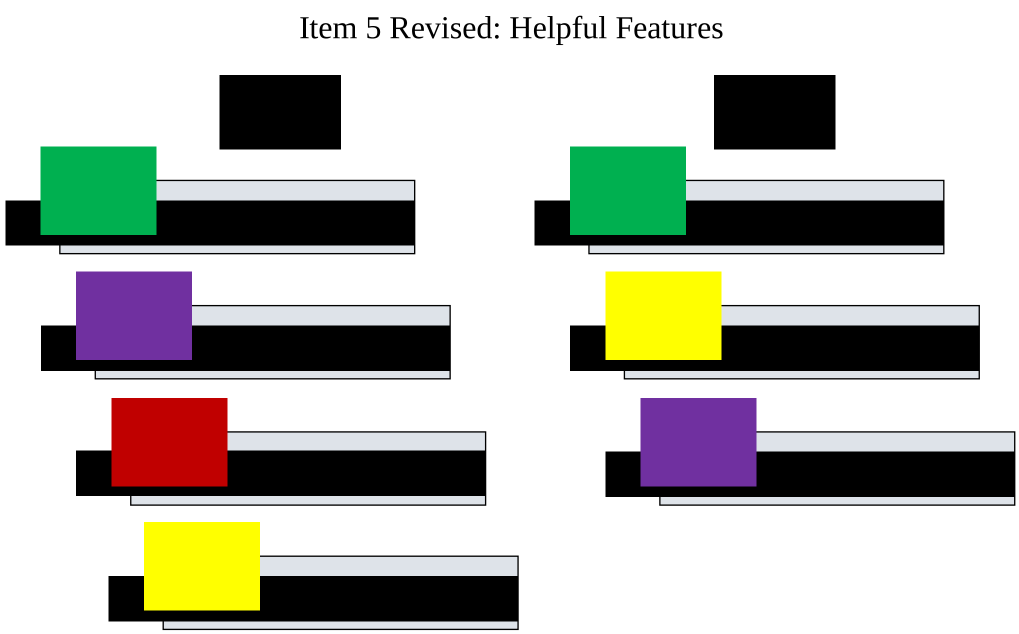


Figure 22. A comparison of the themes for item 5 (revised) helpful features presented in the order of frequency.

These findings suggest that both groups appreciated having a phrase in the item that provided a contextual signal emphasizing their main task in the item. After scaffolding, more ELL students commented on the chemical equation than NES students, who focused more on sentence construction and phrasing. Additionally, ELL students discussed how restructuring the item on the page by adding more spaces between the background information, chemical equations, and question made the item appear more approachable. Consistent with other findings in the current study, the students' responses to this item suggest that NES students rely on language-dependent features more than ELL students, who focus on less language-dependent features, when solving word problems.

ELL students' perceptions of challenging features in item 5 (revised) Although ELL students found many of the modified features of the revised item 5 helpful, they found the item difficult to solve. They discussed several challenging features about this item and made suggestions on how to improve the item. These features are discussed in the order of the following themes: (1) scaffolding, (2) macrostructure, and (3) representation.

Scaffolding. The challenging features identified by ELL students included the mentions of the Haber process and high pressure in the background information of the item. ELL participants discussed that this information was not relevant to be included and that it deterred them from working on the problem as well as it increased their reading time. Sheela mentioned, "The Haber Process, I don't even know what is that…I didn't really need to know that." Lupe stated, "There's not really any areas that is not helpful but the 'high pressure' in the first part, it throws you off […] For this one, you would think of whether you have to use the ideal gas law?

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Another important challenging feature was the added contextualization in the revised version of item 5. ELL students expressed that the storyline was interesting and it was unique to picture themselves be a part of the context of working in a factory; however, reading this information would ultimately take extra time during an exam and did not help them solve the problem. Carlos discussed how having a storyline present in the item was not useful during an exam as it was not tied to solving the question. Carlos expressed, “Because I'm taking the test and I'm just trying to know what the answer is. I do not really visualize myself working in a factory, but it’s okay.”

Macrostructure. ELL students noted that the revised version of item 5 appeared to be more wordy and longer in length than the original version, which was considered challenging. Despite the added spaces between the background information, chemical questions, and question of item, which were identified as helpful features, the item’s overall length made it challenging. The main reason ELL students mentioned for this is because this item requires more reading and interpreting than the original version, which contained fewer words and appeared shorter in length.

Readability. Another problematic feature was what ELL students perceived to be the presence of redundant information. The mention of collecting 700g of ammonia in the last sentence was considered to be repetitive since this information was had already been stated in the first sentence after the chemical equation. Additionally, ELL students discussed that the words “Haber process,” which appeared twice—first in the second sentence and then also in the second sentence were—were redundant. ELL students suggested eliminating the redundant words in order to reduce the overall length of the item. Ammonia was another word that was circled to indicate redundancy as it appears in two sentences near the question of the item. ELL students mentioned that redundant words could be eliminated to reduce their reading time. Sheela commented, “It's more wordy than [item 5 original]. So it has more of a production, which I don't really like it because it's just like waste of time just by reading it when you have a timed exam.”

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Representation. ELL students found the balanced chemical equation to be especially helpful and did not identify challenging features regarding this representation in the item. When asked to suggest further modifications to improve this item, ELL students suggested that adding the chemical formulas of each compound as they are mentioned next to the names would further add valuable representations in the revised item.

NES students’ perceptions of challenging features in item 5 (revised). NES students reported many of the same challenging features that ELL students identified. Overall, NES students generally perceived the revised item to contain irrelevant information which was not effective at providing content support and guidance to solve the problem. The challenging features will be discussed in the following order of emerging themes: (1) scaffolding, (2) readability, and (3) representation.

Scaffolding. NES participants resonated the concerns of ELL students pertaining to background information that is not directly relevant to solving the problem. The revised version of item 5 provides students with information about the Haber process occurring under high pressure. Students considered this information to be problematic as it made them question if this information impacted the reaction in the item somehow. Ana mentioned, "Haber process … it's just extra. Most students probably don't know what that is. We don't need to write, 'in the Haber process,' we could just write, nitrogen gas and hydrogen gas react together to create ammonia gas."

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

The presence of a storyline to add context for students in the item was considered to be unnecessary as it added extra information. Some students mentioned that it is not the type of information they are accustomed to experiencing on chemistry problems. Abby mentioned, "Well, it's kind of like a story but this is chemistry. It's not a story. We can't put ourselves in it. Yes, because we're just trying to figure out the answer."

> **Alex**: I feel when you add more to anything that's circled in red [irrelevant features], that you're just doing it to set a tone and a mood. With only so much time, you're able to have on a test, setting the mood and all that, is unnecessary and no student wants to read it. They're only reading it, because you are making them, because they don't want to miss something.

Readability. NES students also reported that phrases in the question portion contained redundant information. For example, NES students identified that the task of making 700g of ammonia was stated twice since it appears in the sentence before the question itself and then it is restated in the question as shown below. Additionally, the phrase 'you know that this process…' was considered to be confusing as students discussed that it was unclear if the item was asking them to make an assumption or whether they had to perform a computation to prove the yield.

> You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below:
>
> $$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$
>
> You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?

Eric: I don't like it much because it makes me think of algebra questions and geometry questions but I just like more straightforward questions that say that you need to get 70% of ammonia gas or 700 grams of ammonia. How much nitrogen do you need for this?

Macrostructure. When students were asked to compare the revised version to the original version of item 5, they noticed that the revised version contained more words and additional sentences, which signified more information to process. The overall structure of the revised item was reported to appear lengthy, which was considered to be a negative feature for NES students.

Comparison of themes: Challenging features in item 5 (revised). When asked to choose, both ELL and NES students preferred the original version over the revised version of item 5 on an actual exam. Although they appreciated some guided inquiry features embedded in the revised version, both groups preferred the original item for an exam as it contained the least amount of non-essential information. Both groups indicated that the background information was not useful because the process of solving the problem did not require the knowledge of the Haber process of high pressure. ELL students reacted to the longer length and increased number of words in the revised version with concerns about more reading and interpreting time Both ELL and NES students reported that redundancy should be eliminated in the revised version. The results of the emerging themes have been shown in Figure 23.

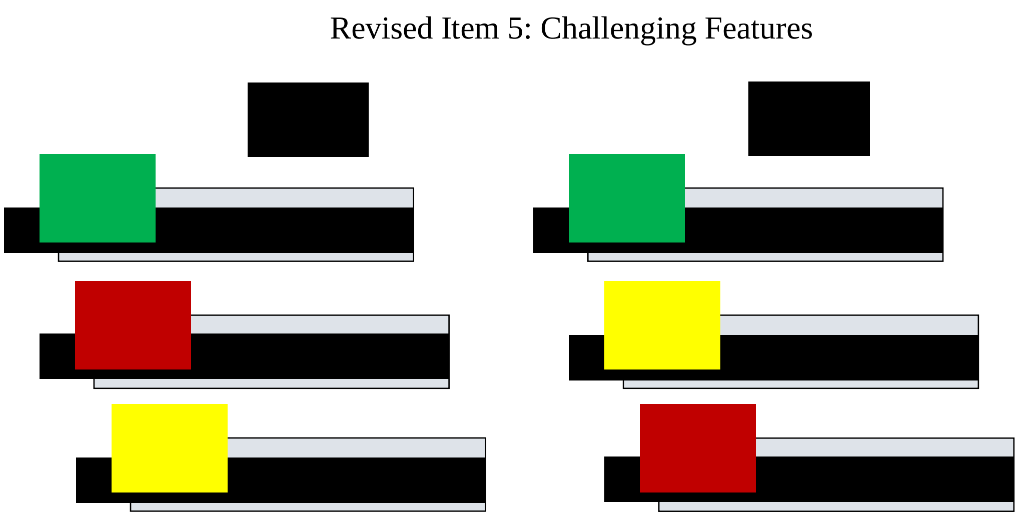


Figure 23. A comparison of the themes for item 5 (revised) challenging features presented in the order of frequency.

The key feature that differentiated revised item 5 from the original version of item 5 was the presence of a storyline that added a real-life context for this problem. Interestingly, both groups of participants did not find the storyline to be a useful feature. Although the storyline provided an opportunity for students to envision themselves as working as a chemist in a factory, participants found it unnecessary and mentioned that it was atypical to see this type of contextual feature on chemistry exam items. Previous studies that employed contextualization by adding a storyline to life science items found that middle school students (both ELL and NES students) responded positively to reading the story within the item (Siegel, 2007; Siegel, 2014). However, this may not be the case for undergraduate students in general chemistry courses, who seem to be more focused on how to numerically solve the problem in a timely manner when taking exams.

Item 6: Original Version

Item 6 focuses on the topic of percent yield. To solve this problem, students are expected to:

1. Use stoichiometry to convert the kilograms of $FeCr_2O_4$ into kilogram of the product, Na_2CrO_4.
2. Recognize that the given numerical value of 1.2 kg of Na_2CrO_4 is the actual yield
3. Calculate the theoretical yield
4. Use that information to compute the percent yield of the reaction.

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

ELL students' perceptions of helpful features in item 6 (original). ELL students paused frequently when reading this item and often made notes on the side of the item while reading through this item during the interviews. In general, ELL students were unable to successfully solve this problem, and most ELL participants could not completely explain how to get to a final answer. When asked to identify helpful features of the item, ELL students' discussed features related to the themes of (1) representation and (2) readability.

Representation. The feature ELL students most frequently reported as helpful in this item was the balanced chemical equation. Students mentioned that they heavily relied on this equation, as the description of the reaction was difficult to understand.

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

ELL students also found the chemical formulas of the compounds appearing next to the name of the compounds to be helpful. Seojun mentioned, "I like how the names and the symbols [are next to each other]."

Readability. ELL students reported that they appreciated the direct sentence structure of the question in the item. For example, the question started with "what is the percent yield…" was considered to be phrased directly and focused the students' attention on the goal of the item, which was to find the percent yield.

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

NES students' perceptions of helpful features in item 6 (original). NES participants generally discussed that this item included a lot of information. For this reason, they needed to spend extra time reading the item. Despite this, most NES students were able to successfully set up this problem and solve it. They discussed the following helpful features related to the following themes: (1) readability and (2) representation.

Readability. NES participants most frequently identified helpful features in the item that related to wording and sentence construction. NES students circled the word "produced," which was an important word to them because it signified that sodium chromate was being made as the product. They also identified the direct phrasing of the question in the item as being helpful. For example, NES students identified the direct phrasing in "what is the percent yield…" as helpful because it emphasized the main goal of the item and allowed them to focus on solving for the percent yield.

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

Representation. NES participants also reported that they found the balanced chemical equation to be essential in this problem as it visually described the entire reaction. The balanced equation showed how chromite and sodium carbonate react to form sodium chromate and other products. It also provided the chemical symbols of each of the compounds described in the background information, which was also perceived to be helpful. The students also appreciated that the equation showed that it had been balanced as indicated by the coefficients in front of each compound.

The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is

$$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$

What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

Comparison of themes: Helpful features in item 6 (original). Both groups identified similar helpful features in original item 6, however, in different order of frequency. ELL students found the balanced chemical equation to be the most helpful part of the item because it provided a visual representation of the chemical reaction, while NES students found the direct sentence structure of the question and the word choice "produce" to be helpful because it helped them understand the item. This finding suggests that ELL students rely more on language independent features in accessing a word problem, such as representation; where as NES students utilize more language-dependent features of the world problem, such as sentence construction and word choices. A comparison of these themes is shown in Figure 24.

Item 6 Original: Helpful Features

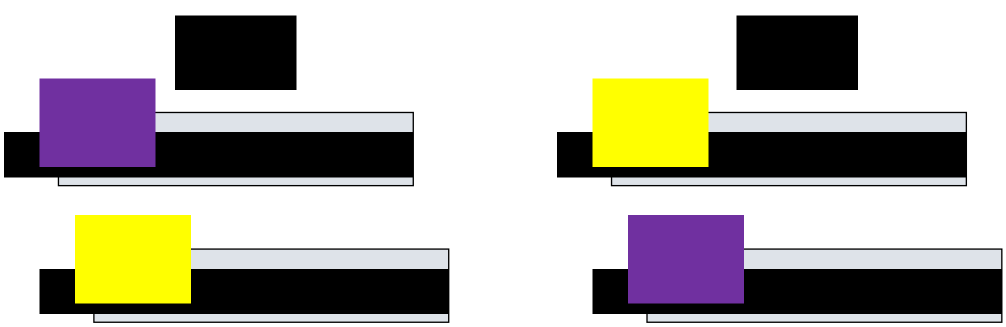


Figure 24. A comparison of the themes for item 6 (original) helpful features presented in the order of frequency.

An important point to note here is that despite both groups identifying similar helpful features, ELL students were unable to successfully set up this problem. This is problematic because it indicates that the identified helpful features related to representation and readability were not enough to support ELL students' performance in this item. According to literature that addresses fairness and validity issues in assessments, an item such as this could be detected as one that carries differential validity for different groups of test takers and should be reviewed for construct irrelevance variance (Abedi et al., 2004; Abedi, 2007; Haladyna & Downing, 2004; Turkan & Liu, 2012).

ELL students' perceptions of challenging features in item 6 (original). ELL participants discussed many difficulties with interpreting this item. They identified several features that made understanding this item challenging and offered suggestions on how to improve these features. The reported challenging features are discussed in the following order of emerging themes: (1) readability, (2) scaffolding, and (3) representation.

Readability. ELL students predominantly perceived that it was difficult to comprehend the item because of word choices and complex sentence structures. The word "roasting" was not one that ELL students were familiar with, and they perceived the use of it to be confusing because they questioned how it was involved in the problem. They also mentioned that it was hard to follow the first sentence and, as a consequence they had to re-read the first sentence many times to get a sense of what was being communicated about chromite.

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

Rohan: I don't understand how [roasting] would help me with the question. It's just a weird word…because when you think of roasting you think of like heat. Then, you had to think about if heat is lost or gained in the equation. So, it might confuse people, it's just not a great word to put in there.

Additionally, ELL students also mentioned that the latter part of the question was difficult to comprehend as it contained a lot of information in one long, complex sentence. Although the sentence began with a direct question, ELL students discussed that the sentence's syntax made it confusing because it included numerical information of two different compounds in the same phrase. They suggested separating this complex sentence into simpler sentences.

Ina: I know most chemistry equations are written like this. It's just when there are two numbers back to back like this, I don't know which one to use first on the conversion factor thing. Yes, it's just a bit confusing.

Scaffolding. ELL students expressed their concerns about the usefulness of the background information and the lack of content support in the item. ELL students found that the description about the making of chromite was not useful in solving the problem. When asked how they would further modify this item to improve it, ELL students suggested that the first sentence should be eliminated as this information did not provide support for solving the problem. Additionally, ELL students mentioned that this item was difficult to solve because there was a lot of information embedded in the question portion of the item and there was a lack of guidance on how to begin setting up the problem.

The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is

$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$

What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

Representation. ELL students questioned why the chemical formula for chromite was missing when the chemical formulas for sodium chromate and sodium carbonate were given. Students felt that it was inconsistent to not also have the chemical formula for chromite. Eventually, they were able to figure out that chromite was $FeCr_2O_4$ based on the chemical equation provided; however, they expressed uncertainty that this was the correct compound for chromite and asked me to verify.

The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is

$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$

What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

NES students' perceptions of challenging features in item 6 (original). NES participants generally expressed their frustration about the way this item was written. The challenging features NES students reported are reported in the following order of emerging themes: (1) readability, and (2) scaffolding.

Readability. The most frequently mentioned challenging features for this item related to its sentence structure in the first and last sentences, and the use of the word "roasting." NES students discussed that the question portion of the item is unclear. It was difficult to determine that they were being asked to look for the percent yield of sodium chromate because the sentence structure contained too much information.

> **Josh**: You put two values in the same question, but they are not supposed to be used in the same step, that gets confusing. Even if it was just 1.2 kilogram of sodium chromate was produced, period. What is the percent yield from that or contains -- just separating those [values] even in the slightest bit. Because when you put in the same question, an average student wouldn't probably know [how to start].

The sentence construction in the first sentence was also reported to be challenging, especially due the overuse of the word "chromium."

> **Jasper** mentioned: I think it's very wordy or dense, especially in the beginning. I read through and it said, 'Use a source of chromium in chromium compounds,' they really just put chromium and chromium back together, but because compound was [shown below], I said, 'Oh, okay.'

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

Scaffolding. Another challenging aspect of the original item 6 indicated by NES students was the irrelevant background information and lack of content support. NES discussed that the background information provided that described chromite was unnecessary and should be removed. Students conveyed that this information was not only confusing to understand, it also did not serve a purpose in helping them setup the problem. NES students suggested that the item should include more content support in the form of guided steps to help students begin setting up the problem.

> The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is
> $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$
> What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$?

Comparison of themes: Challenging features in item 6 (original). Both ELL and NES students found the original version of item 6 to be a difficult one to follow, mainly because of confusing wording and sentence structures. The lack of content support was a concern for both groups of participants. ELL students suggested that adding the chemical formula for chromite would have helped them follow the reaction better. A comparison of the themes that encapsulate the challenging features between ELL and NES participants is shown in Figure 25.

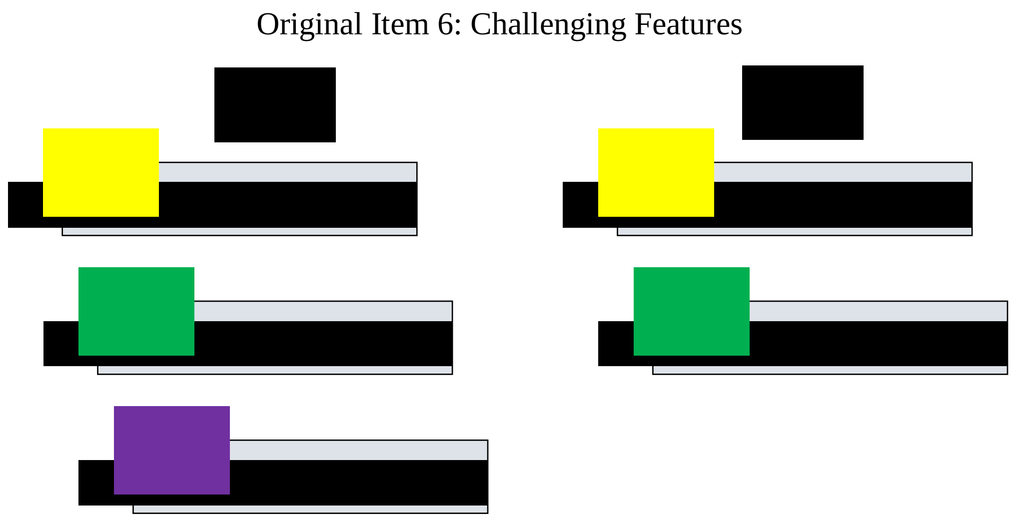


Figure 25. A comparison of the themes for item 6 (original) challenging features presented in the order of frequency.

These findings suggest the challenging features impacted ELL students more than NES students as ELL students were unable to set up the item successfully. Readability-related issues were identified by both groups as challenging; however, NES students were still able to interpret this item despite the presence of confusing wording and sentence structures. As noted previously, one potential explanation for this finding is offered in literature, which suggests that there could be linguistically complex elements embedded in item that differentially favor NES students' performance, which attribute to the performance gap between ELL and NES students (Abedi, 2007; Abedi et al., 2004; Haladyna & Downing, 2004; Turkan & Liu, 2012).

Item 6: Revised Version

The original version of item 6 was revised according to the EFCA guidelines as discussed in Chapter 3 and item modifications listed in Appendix A. The background information was restructured to reduce non-essential information. Chemical formulas for all compounds discussed in the background information were added next to the names of each compound. Physical spaces were added to separate the background, balanced equation and the main question portions. The sentence construction of the question in the item was also modified to clarify that students should calculate the percent yield of Na_2CrO_4.

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite, $FeCr_2O_4$, with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

ELL students' perceptions of helpful features for item 6 (revised). In general, ELL students found the aforementioned modifications to be helpful in understanding the item. More ELL students were more able to successfully set up the problem and explain how they would get to the final answer with the revised version than the original version of item 6. They identified helpful features related to the following themes: (1) readability, and (2) representation.

Readability. ELL students indicated that it was helpful to read fewer words in the revised version of the item. They also noted that the simpler sentence structure was easy to comprehend and that specific word choices, such as "made from" instead of "produced," were easier to follow. Lupe mentioned, "This one says 'made from' instead of 'produced,' which is better." Naima indicated that the "[Revised version was] easier to read and understand the question…you don't get overwhelmed by so many parts."

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite, $FeCr_2O_4$, with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

Representation. Another aspect that ELL students reported to be helpful in the revised version was having the chemical formulas of compounds presented next to the name of each compound. This was helpful because the chemical formula of chromite was shown as it was being discussed, unlike the original item, which seemed to elicit uncertainty among ELL students. The addition of the chemical formulas next to the name of each compound made it easier to follow the description of the reaction.

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite, $FeCr_2O_4$, with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

NES students' perceptions of helpful features for item 6 (revised). NES participants generally found the revised version the item 6 to be easier to solve and follow than the original version. They indicated helpful features related to the following themes: (1) readability, and (2) representation.

Readability. NES participants indicated that the revised version of item 6 was easier to read than the original version because it was less wordy, contained shorter sentence structures and fewer redundant words. These helpful elements made the item more explicit for NES students. Specifically, the second sentence of the item discusses the mixing of the reactants, chromite and sodium carbonate, and then states that the product, sodium chromate, is made. NES students indicated that this sentence was easier to follow than the corresponding sentence in the original version.

Eric: It's more explicit in what it's doing. It's like you mix this and this, and you get this. Unlike the other [original] one, it says that, but it's not as wordy about it. Yes, that's why I think this one is probably easier.

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite, $FeCr_2O_4$, with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

Representation. NES participants also noted that having the chemical formulas next to the name of each compound was another helpful feature in the item in addition to the balanced equation provided. Having chemical formulas values next to the name of each compound as they are discussed helped students make the correct associations with the compounds and their symbols. The balanced chemical equation was effective at illustrating the overall reaction that took place to make Na_2CrO_4. Alex mentioned, "The chemical formula is okay, but then having the names by it does make it a little easier."

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite, $FeCr_2O_4$, with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

Comparison of themes: Helpful features for item 6 (revised). Both ELL and NES participants found the modifications in the revised version of item 6 to be helpful and both groups indicated that they would prefer to see the revised version on an exam. It is interesting to note that both groups of participants identified similar features as helpful in this item. There were no major differences between the perceptions of ELL and NES students in terms of the features identified as well as the order of the emerging themes for this item. This finding suggests that the modifications employed in this item, as recommended by the guidelines of the EFCA (discussed in Chapter 3 and explained in Appendix B), served to support both groups of participants by reducing the cognitive load and making the item more accessible. A comparison of themes is shown in Figure 26.

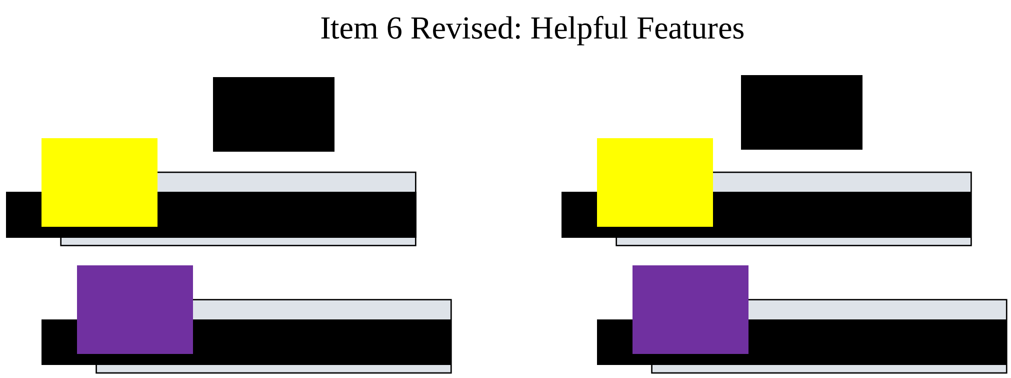


Figure 26. A comparison of the themes for item 6 (revised) helpful features presented in the order of frequency.

ELL students' perceptions of challenging features for item 6 (revised). Although ELL students found the revised version to be easier to understand the original, they were able to identify several challenging features in it that could be improved. To that end, ELL students identified challenging features related to (1) readability and (2) scaffolding.

Readability. Although the sentence structure in the main question was revised, ELL students reported that there was still too much information presented in the question, which is in the last statement of the item. This statement (shown below) was perceived to be confusing and overwhelming to interpret. ELL students suggested that multiple compounds with multiple numerical values not be presented in one, complex sentence.

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

Scaffolding. Another challenging aspect of the revised version of item 6 identified by ELL students was related to what they saw as irrelevant information included in the item. ELL students conveyed that the background information provided was not useful and did not provide cues to help set up the problem. In particular, the sentence regarding the description of chromite was not thought to be helpful. Sheela expressed, “[Background information] Kind of and kind of not [helpful] because the beginning, it wants to be helpful but it's not really helpful. Because it's saying, chromite is the source of chromium and chromium compounds.”

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

NES students’ perceptions of challenging features in item 6 (revised). Although NES students preferred the revised version over the original version of item 6, they indicated that it included confusing features related to (1) scaffolding and (2) readability.

Scaffolding. The biggest challenge reported by NES students regarding the revised version of item 6 related to the background information about chromite, which was considered to be unnecessary. Students found the additional information about chromite was not only irrelevant, but also confusing because it made them question if "chromium and chromium containing compounds" needed to be somehow used in the chemical equation. For example, students questioned if they needed to look for other "chromium containing compounds" in the products of the reaction. Eventually, they were able to figure out that this information was not needed to solve the problem.

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

Readability. Another challenging feature indicated by NES students was the sentence structure in the question portion of the item. NES students had a similar perception as ELL students in that they reported that it was confusing to read about multiple numerical values and compounds in the same phrase. During the interview, they circled the part after the percent yield of Na_2CrO_4 and indicated that there was too much information here. It was suggested to this question could be simplified by dividing its information among multiple sentences, but did not provide concrete examples of how to do so.

> Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation:
>
> $$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$
>
> What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?

Comparison of themes: Challenging Features for item 6 (revised). Both ELL and NES students perceived that the background information provided in the revised version of item 6 was not relevant to solving the problem and that the main question part of the item was constructed in a convoluted manner. Both groups of students made similar suggestions to improve these features of the item. Despite reporting these challenging features, ELL and NES students were able to set up the problem successfully. A comparison of the emerging themes between ELL and NES participants is shown in Figure 27.

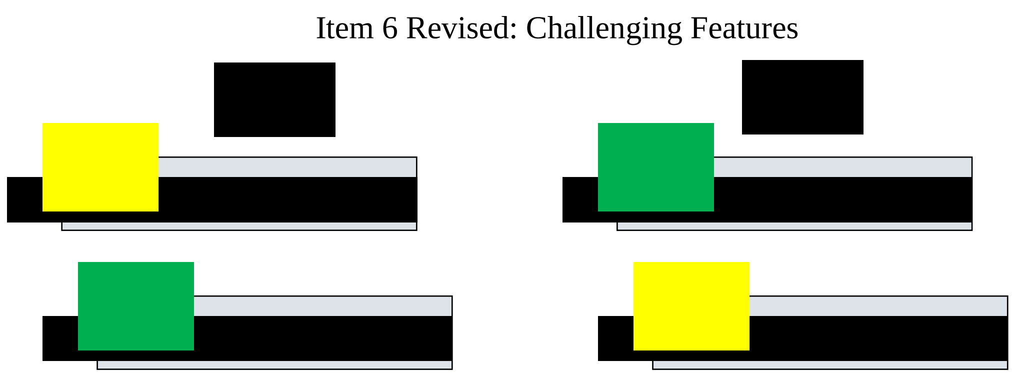


Figure 27. A comparison of the themes for item 6 (revised) challenging features presented in the order of frequency.

This finding suggests that when there are multiple pieces of information provided in one sentence on assessment items, both groups of students struggle with interpreting the information. Both groups' participants conveyed they experienced cognitive overload when reading these types of statements; however, more ELL students indicated that this feature was particularly challenging for them than NES students. This suggests that ELL students are more prone to experiencing difficulty on complex readability-related features than NES students as affirmed in literature (Noble et al., 2014).

Chapter Summary

In this chapter, I examined students' perceptions of four original and revised assessment items and their suggestions about ways to improve the item. I found that both groups of students appreciated most of the modifications of the EFCA applied in the revised versions of the items, including linguistic simplification, division of prompts, and addition of representations. The added background information was generally considered unnecessary and cumbersome by both groups of participants. Accordingly, both groups of students preferred the revised versions of items 1, 2, and 6 on an actual exam.

The item 5 was the only revised item presented to participants which was not preferred by students. Contextualization in the form of a storyline was a key feature included in the revised item 5, which the students did not perceive to be as helpful as was intended by the EFCA. The added context in the revised item 5 provided a storyline that placed the student in a factory that produces ammonia. Both groups of students found the storyline in this item to be unnecessary and preferred the original version of item 5. This finding is not in alignment with previous research that reported that storylines in written assessment items for middle school ELL and NES students for life sciences functioned was valuable tool to engage the students in the item and elicit learning (Siegel et al., 2014). The results of the current study described that both groups of undergraduate students did not find that the storyline feature played a role in solving the problem, which made it unnecessary and problematic as it was wordier and required more reading time than the original item.

In the same vein, the findings suggest that undergraduate students' expectations of exam items played a significant role in how they responded to specific features in the items. The results revealed that both groups of participants only preferred to see essential information that directly contributed to solving the problem. Although some students mentioned that the background information did provide context and was interesting, their responses implied that "fun facts" and/or "cool information" should be saved for homework or in-class practice problems. This suggests that undergraduates have been trained to view chemistry exam items as purely a means to solve the problem to get to a correct answer in the shortest amount of time as opposed to an opportunity to learn, be engaged and demonstrate their knowledge. One of the principles of the EFCA (discussed in Chapter 3) denotes that assessments should elicit learning; however, students' perceptions reveal they are not accustomed to viewing the test as a learning opportunity.

Overall, the features identified by students in the assessment items reveal significant implications about how students perceive assessment items in general chemistry. The finding that both groups of students reacted similarly to the modifications embedded in revised versions of the items suggest that the EFCA maybe a useful tool to revise preexisting written assessment items in a way that is more accessible for all students. The implications of these findings are discussed in the final chapter.

CHAPTER 6 CONCLUSIONS & IMPLICATIONS

Project Summary

This study attempted to modify cognitively difficult features of written assessment items on the topics of limiting reactant and percent yield in general chemistry courses to be more accessible for undergraduate ELL students. The goal of this study was to make the assessment items more accessible for ELL students; however, the end result was that all students saw the items as more accessible. The modifications used to refine the items were derived from the EFCA, which included simplification of vocabulary and syntax, reduction of non-essential information, and addition of visual support. Students were shown either the original (unchanged) version or the revised version of the assessment item in random order. During semi-structured interviews, students were asked to discuss the following: what the item was asking, and features of the item that were helpful and/or challenging, and how they would change the question to further improve it.

There were several interesting findings in the current study. First, ELL and NES students utilized different types of features to develop an understanding of and set up the problems in the items. ELL students largely used language-independent features to make meaning and set up the items while NES students used more language-dependent features to understand and set up the items. Second, features related to macrostructure played an important role in changing the cognitive load of an item for both groups of students. Third, compared to NES students, more ELL students indicated that they found scaffolding in the form of content support and guidance within items to be helpful in solving the problems. Finally, neither ELL students nor NES students found features that provided context for the problems to be useful when the context was not directly related to solving the problem. In fact, students suggested that this type of information be eliminated on exam items.

The Equity Framework of Classroom Assessment

A significant goal of this study was to examine the effectiveness of the extending the EFCA in written assessment questions in general chemistry through the voices of students. In previous studies, the EFCA approach to scaffolding life science assessment items for middle school ELL students yielded promising findings (Siegel, 2007; Siegel et al., 2014). The results of the current study suggest that the EFCA is applicable to general chemistry assessment items and beneficial overall to undergraduate students; however, not every EFCA modification utilized in the items was perceived to be helpful by students.

Of the 11 modifications originally proposed in the EFCA, the four assessment items used in this study included eight. A list of these modifications alongside the items that utilize each them as well as whether each modification was perceived to be helpful or challenging is presented in the table below.

Table 2. Summary of how students perceived each modification.

Modifications	Items				Helpful	Challenging
	1	2	5	6		
Simplification of vocabulary and syntax	x	x	x	x	x	
Replacement of sentences into lists	x				x	
Reduction of non-essential information	x		x	x	x	
Addition of visual support	x	x			x	
Rearrangement of data		x	x	x	x	
Addition of active voice and direct sentences		x	x		x	
Division of prompt into smaller units		x	x		x	
Contextualization (adding a storyline)		x	x			x

As shown in Table 2, contextualization was reported by both groups of students as a challenging feature for the purposes of the exam; however, students may respond differently if contextualization were included on homework items, where time to complete word problems is less of an issue. Contextualization was a dominant feature employed in revised item 5, which added a storyline to the problem by telling students that they work in a factory. Based on the principles of the EFCA, adding context to science assessment items helps ELL students find cues and may help them interpret the academic language embedded in the item (Siegel et al., 2008). In the case of item 5, the added storyline did not provide enough scaffolding to help students identify and set up the problem at hand, which was to calculate the amount of nitrogen needed to make ammonia. Many students expressed that the storyline was an interesting feature because most chemistry questions are not typically written this way; however, the storyline did not provide the support needed to make the item more accessible. Additionally, many ELL participants expressed that it made the item lengthier, which would add to their reading time on an actual exam.

On the other hand, participants found other modifications implemented in the revised versions to be helpful. ELL students, particularly, found that scaffolding-related features–such as division of prompts into smaller units that provided guidance—to be useful, especially in solving multi-step problems with difficult readability. From the perspective of developing language proficiency, scaffolding has been shown to help students identify key parts in an item and to provide content support to solve the problem (Nobel et al., 2012; Siegel, 2007). In the current study, ELL students' responses suggest that features that guided their understanding and supported their ability to make connections across tasks alleviated some of the cognitive demands of the item.

Conclusions

The findings suggest that the principles of the EFCA are applicable to general chemistry assessment items and beneficial to both ELL and NES undergraduate students. An important part of using the EFCA for modifying exam questions was reconstructing the item in a way that is accessible for students who are not fully fluent in the English language while maintaining the content difficulty. By removing linguistically complex elements, adding visual clarity and guided inquiry into original exams items, challenging features consisting of technical vocabulary and content specific academic language were mitigated. Students were better equipped to solve the word problems.

Another significant goal of this study is to promote fairness in classroom assessments for all students, particularly for students who are still learning the English language. In this study, ELL students especially benefitted from the modifications applied to exam items. Most ELL students were unable to successfully explain their understanding of many of the original items. If these items were shown on an actual course exam, many ELL students would be unable to complete their work and, as a results, would have performed poorly. In a typical college chemistry course, their lower exam scores could be thought of as a result of lack of studying, dedication, intelligence, etc.; however, the results of this study suggest that the test score gap between ELL and NES students could be reduced by modifying the exam items. The responses of ELL students indicated that revising the items mitigated linguistically complex features and they were able to better understand, explain, set up and solve the problems.

Interestingly, participants whose first language is English also perceived most of the item modifications positively. NES students showed similar preferences as ELL students in choosing revised items and reiterated similar perceptions of helpful and challenging features in all items. Although more NES students were able to successfully set up the original versions of the questions than ELL students, they expressed that certain parts of the original items were still challenging to follow and should be improved. When presented with the revised versions, NES participants were able to explain the items more confidently. This finding suggests that revising general chemistry exam questions can benefit not only ELL students, but all students in a general chemistry course.

This study explores the complex issue of classroom assessments at the postsecondary level. Research that focuses on the needs of linguistically underrepresented university students in science courses is rare, but especially valuable in understanding the nature of testing in college classrooms. Given the constraints of teaching a historically difficult subject at the university setting, this study informs the development of more equitable assessments in college chemistry courses. It also acts as a foundation for future studies that focus on changing classroom assessments in a way that benefits the learning needs of students with developing English language proficiencies as well as NES students in general chemistry.

Broader Implications for Teaching

The results of the current study have implications for teaching and learning. Assessing students with limited English language proficiencies is a tricky issue; however, as educators, we have an obligation to give all students an equal opportunity to demonstrate their knowledge by minimizing sources of measurement error. The results of this study reveal that the EFCA can be applied to assessment items in general chemistry and this approach may represent a more promising approach for supporting ELL students in the college classroom settings than using test accommodations (e.g., extra time, dictionaries, translators, etc.). Designing equitable assessments will require the careful consideration and examination of many factors including, but not limited to, the role of language and culture of students, scientific understanding, and effective assessment practices. The results of this study will need to be verified through future studies that focus on different topics in general chemistry to further confirm the effectiveness of the EFCA modifications for making chemistry classroom assessments more accessible and equitable for all students.

Limitations

Conducting this qualitative study with undergraduate students in general chemistry led to new discoveries as well as limitations that must be mentioned. The ELL student population recruited for this study included ELLs who have been residing in the U.S. for 10 years or less.

This group of students exhibited a wide range of English language proficiencies from students who were advanced English learners to students who were beginners. Narrowing this time frame range and differentiating the levels of English proficiencies would have yielded a more holistic understanding of how different ELL students interpreted the assessment items. Also, it is important to mention that the ELL students in this study had different ethnic and language backgrounds, which may have played a role in how they approach assessment items in general chemistry.

Another limitation of the current study is that it was not implemented in a specific general chemistry course. For the purposes of the study, undergraduate students were recruited from the general chemistry I and general chemistry II courses. Many participants had different chemistry instructors, who had different teaching and testing styles, which may have impacted the way students learned the topics of limiting reactant and percent yield.

Future Research

Future research is needed to inform the field and contribute to the body of literature on equitable assessments. As an extension to the current study that examined ELL and NES students' perceptions of features of the assessment items qualitatively, a future study that measures students' performance on original versus revised exam questions quantitatively would beneficial in understanding the effects of revising assessment items on student exam scores. It would also be helpful to understand how this framework can be applied to different chemistry topics. Additionally, future work should also investigate the perceptions of ELL students who have the same language backgrounds as this may play an important role in their perceptions of assessments.

One of the findings of the current study suggests that undergraduate students' mindsets about exam items compared to other non-exam items, such as homework or in-class practice word problems, are different. Their perceptions of helpful and challenging features were largely guided by the fact the items presented to them were exam items. An interesting future study would be focused on understanding how students think of homework word problems compared to exam problems in terms of the usefulness of the EFCA modifications implemented.

This study suggested that students hold their own expectations of exam items; however, these expectations may be different from instructors' expectations of exam items. A future study could focus on examining how general chemistry instructors think of assessments in their courses. Understanding the methods instructors are currently using to design their classroom assessments, and the type of exam items they find to be useful for assessing students' content knowledge will add a much needed dimension to this body of research.

Another critical direction for future research focuses on the beliefs of faculty who teach general chemistry. Understanding the instructors' perspectives about evaluating students of linguistically diverse background in chemistry is a key aspect of advancing the field of equitable assessments. Most studies suggest that teachers are unaware of ways to teach and assess ELL students, especially in the context of science (Lee, 2005). However, it is important to understand how much instructors know about the needs of ELL students in the classroom. The bulk of research about how instructors approach teaching culturally and linguistic minority students has centered on K-6 classrooms (Buxton et al., 2008; Cho & McDonnough, 2009; Lee, 2004; Swanson et al., 2014); however, research on understanding this issue at the postsecondary level is scarce but necessary.

APPENDIX A

List of Item Modifications

1. Linguistic simplification of vocabulary and syntax:
    a. Linguistic simplification of vocabulary: removing unnecessarily complex words and/or phrases for simple terms that convey the same meaning
    b. Linguistic simplification of syntax: Replacing long sentences with embedded commas and/or semicolons with shorter and more direct statements that convey the same information in terms of content
2. Replacement of sentences with lists
    a. Modifying sentences that give more than one piece of information and adding a bulleted list to separate the pieces of information
3. Reduction of nonessential information
    a. Reducing the *number of words* in the item by removing unnecessary words that add to the overall reading time for students. This also includes removing extraneous words that are not necessary to understand and solve the problem, and removing content information that is irrelevant to solving the problem
4. Addition of visual supports in the stem of item
    a. Making the information in the item more visually accessible by formatting the question in a way that easily differentiates background information in the question stem (portion that gives background information) from pertinent information in the question. For example, adding paragraphs, line breaks, adding/removing space between background information, etc.

5. Division of data:

    a. Rearranging the order in which the information appears in the question so that it is logical and easy to follow.

    b. Adding a data table that contains critical information relevant to the problem

6. Alignment with the language of instruction

    a. Monitoring the level of academic language used during instruction and matching the level of vocabulary in exam items. This would be outside the scope of my study since I will not be attending participants' classrooms. Therefore, I have not included an example of this modification.

7. Alignment of the language within the item

    a. Ensuring that the tense, voice, and overall structure of the item are consistent.

8. Use of bold type for emphasis

    a. Highlight an important phrase or word(s) in the item that is crucial for solving the problem

9. Addition of graphic organizers in the prompts

    a. Adding illustrations and/or representations in order to better describe the problem

10. Division of the prompts into smaller units

    a. Dividing the question prompt into smaller, more comprehensible units if the question prompt is more than three sentences

11. Contextualization of the test item

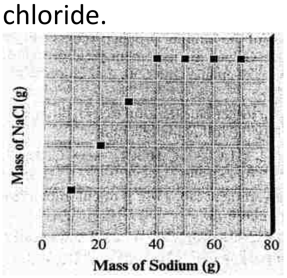
    a. Adding meaning by embedding contextual cues that help stage the test item as a problem

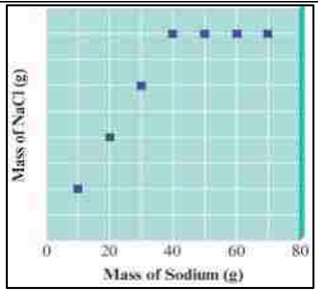
    b. Making connections between parts of the questions to establish flow

c. Adding more steps to scaffold all parts of the problem

APPENDIX B

Chemistry Questions

Original	Revised	Modifications Applied
1. Methanol (CH_3OH), also called methyl alcohol, is considered to be the simplest alcohol. It is used as a source of fuel in race cars and is a potential replacement for gasoline. Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen. Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$. Calculate the theoretical yield of methanol. If 3.57 x 10^4g CH_3OH is actually produced, what is the percent yield of methanol? <u>Source:</u> Zumdahl, S.S. & Zumdahl, S.A. (2008) *Chemistry* (8th ed.). Belmont, CA: Cengage Learning. (p. 113).	Methanol (chemical formula: CH_3OH) is the simplest alcohol. It can replace gasoline, and it is used as fuel in race cars. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2). Knowing that methanol is made when you react CO with H_2, what would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$? You found that that 3.57x10^4g CH_3OH is actually produced. What is the percent yield of methanol?	1: Linguistic simplification of vocabulary and syntax 2: Replacing sentences with lists 3: Reduction of words in the item stem 4: Adding to visual support in the stem of items
2. You have seven closed containers, each with equal masses of chlorine gas (Cl_2). You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on. Sodium and chlorine react to form sodium chloride.  Explain the shape of the graph. Calculate the mass of NaCl formed when 20.0g of sodium is used. Calculate the mass of Cl_2 in each container. Calculate the mass of NaCl formed when 50.0g of sodium is used. Identify the remaining reactant,	There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas ($Cl_{2(g)}$) in it. You add sodium ($Na_{(s)}$) as follows: Jar 1: 10.0g $Na_{(s)}$ Jar 2: 20.0g $Na_{(s)}$ Jar 3: 30.0g $Na_{(s)}$ Jar 4: 40.0g $Na_{(s)}$ Jar 5: 50.0g $Na_{(s)}$ Jar 6: 60.0g $Na_{(s)}$ Jar 7: 70.0g $Na_{(s)}$ Sodium and chlorine react to make sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data.	1: Linguistic simplification of vocabulary and syntax 4: Addition of visual supports in the stem of item 5: Division of data 7: Matching the language within the item more precisely 9: Addition of graphic organizer in prompt 10: Division of prompts into smaller units 11: Contextual scaffolding

<table>
<tr>
<td>and determine its mass for parts b and d above.
Source:
Zumdahl, S.S. & Zumdahl, S.A. (2012) Chemistry: An atoms first approach (2nd ed.). Boston, MA: Cengage Learning. (p. 241).</td>
<td>

Answer the following questions about this reaction and the data in the graph:

A. Write the balanced equation for the reaction between Cl_2 gas and sodium.
B. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in Part A, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2?
C. Based on the amount of NaCl that you calculated in Part B, label the y-axis of your graph.
D. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?
E. Use the balanced equation from Part A to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5.
F. How does the amount of Cl_2 you calculated in Part E compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?
G. What is the mass of the leftover reactant in jar 5?</td>
<td></td>
</tr>
</table>

	H. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following: I. Why the amount of NaCl produced increases at first when you add more sodium? J. Why is the amount of NaCl produced constant when you add more than 40g of sodium?	
3. When ethane (C_2H_6) reacts with chlorine (Cl_2), the main product is C_2H_5Cl, but other products containing Cl, such as $C_2H_4Cl_2$, are also obtained in small quantities. The formation of these other products reduces the yield of C_2H_5Cl. Calculate the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2, assuming that C_2H_6 and Cl_2 react only to form C_2H_5Cl and HCl. Calculate the percent yield of C_2H_5Cl if the reaction produces 206 g C_2H_5Cl. Source: Brown, T.L., LeMay, H. E., Bursten, B.E., Murphy, C.J., Woodward, P.M. & Stoltzfus, M.W. (2015). *Chemistry: The central science* (13 ed.). Upper Saddle River, NJ: Pearson (p. 118).	When ethane (C_2H_6) reacts with chlorine (Cl_2), the reaction produces C_2H_5Cl as the main product. Other products containing chlorine are also produced in small amounts. These are called minor products. Minor products decrease the amount of C_2H_5Cl (main product) that is made in this reaction. A. Write the balanced equation for the reaction between ethane and chlorine that produces C_2H_5Cl and HCl as products. B. Using the balanced equation you wrote for Part A, find the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2? C. If you conducted the reaction described in Part B in the laboratory and only collected 206g of C_2H_5Cl, what is the percent yield of C_2H_5Cl?	1: Linguistic simplification of vocabulary and syntax 2: Replacing sentences with lists 3: Reduction of nonessential information 4: Addition of visual support in the stem of item 5: Division of data 10: Division of the prompts into smaller units 11: Contextual scaffolding
4. When hydrogen sulfide gas is bubbled into a solution of sodium hydroxide, the reaction forms sodium sulfide and water. How many grams of sodium sulfide are formed if 1.25g of hydrogen sulfide is bubbled into a solution	When hydrogen sulfide gas is mixed into a solution of sodium hydroxide, the reaction makes sodium sulfide and water. You are told that this reaction gives you a 92.0% yield for sodium sulfide when you do the reaction in the	1: Linguistic simplification of vocabulary and syntax 3: Reduction of nonessential information 5: Division of data 10: Division of prompts into smaller units

containing 2.00g of sodium hydroxide, assuming that the sodium sulfide is made in a 92.0% yield? Brown, T.L., LeMay, H. E., Bursten, B.E., Murphy, C.J., Woodward, P.M. & Stoltzfus, M.W. (2015). *Chemistry: The central science* (13 ed.). Upper Saddle River, NJ: Pearson (p. 118).	laboratory. How many grams of sodium sulfide will you make in the laboratory if you mix 1.25g of hydrogen sulfide with 2.00g of sodium hydroxide?	11: Contextual scaffolding
5. The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following: $N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$ If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%? Adapted from: Brown, T.L., LeMay, H. E., Bursten, B.E., Murphy, C.J., Woodward, P.M. & Stoltzfus, M.W. (2015). *Chemistry: The central science* (13 ed.). Upper Saddle River, NJ: Pearson (p. 338).	You work in a factory that manufactures ammonia gas. The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below: $N_2(g) + 3H_2(g) \rightarrow 2NH_3(g)$ You know that this process will give you a percent yield of 70% for ammonia gas. Your job is to make 700g of ammonia. What mass of nitrogen do you need to use so you can collect 700g of ammonia?	1: Linguistic simplification of vocabulary and syntax 5: Division of data 7: Matching the language within the item more precisely 11: Contextual scaffolding
6. The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite. For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow$ $8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$ What is the percent yield if 1.2kg of Na_2CrO_4 is produced from ore that contains 1.0kg of $FeCr_2O_4$? Source: Bishop, M. (2001). An Introduction to Chemistry. Chiral Publishing Company.	Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. This reaction is shown below in the balanced equation: $4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow$ $8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$ What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$?	1: Linguistic simplification of vocabulary and syntax 3: Reduction of nonessential information 5: Division of data 10: Division of prompts into smaller units

7. How many grams of calcium hydroxide are needed to make 3.75 grams of calcium phosphate? What if a mistake was made and we used 2.37 grams of $Ca(OH)_2$ and 2.69 grams of H_3PO_4? Which compound is the limiting reagent? How much $Ca_3(PO_4)_2$ would be obtained? What if 2.37 grams of $Ca(OH)_2$ were used along with excess H_3PO_4 and only 2.98 grams of $Ca_3(PO_4)_2$ were obtained (instead of the theoretical yield)? What would be the percent yield? Source: Instructor's test bank	Calcium hydroxide ($Ca(OH)_2$) and phosphoric acid (H_3PO_4) react to produce calcium phosphate ($Ca_3(PO_4)_2$) and water. Write the balanced equation for this reaction. How many grams of calcium hydroxide ($Ca(OH)_2$) do we need to measure out to make 3.75g of calcium phosphate ($Ca_3(PO_4)_2$)? We made a mistake in our measurements. Instead of using the amount of calcium hydroxide you calculated in Part B, we used 2.37g of $Ca(OH)_2$ and 2.69g of H_3PO_4. How many grams of $Ca_3(PO_4)_2$ should be made under these conditions? Which compound is the limiting reagent for the reaction described in Part C? What would be the percent yield if we only collected 2.98g of calcium phosphate when we did the reaction described in Part C in a laboratory?	1: Linguistic simplification of vocabulary and syntax 3: Reduction of nonessential information 4: Addition of visual support in the stem of item 7: Matching the language within the item more precisely 9: Addition of graphic organizer in prompt (adding chemical formulas after the name of the compound) 10: Division of prompts into smaller units 11: Contextual scaffolding

APPENDIX C

Item Modification Procedure

Question 1

~~Methanol (CH_3OH),~~ ~~also called methyl alcohol~~, ~~is considered to be the simplest alcohol~~. Methanol (chemical formula: CH_3OH) is the simplest alcohol **(1)**. ~~It is used as a source of fuel in race cars and is a potential replacement for gasoline~~. It can replace gasoline, and it is used as fuel in race cars. **(1, 3)** ~~Methanol can be manufactured by combining gaseous carbon monoxide and hydrogen~~. Methanol is made by mixing gaseous carbon monoxide (CO) and gaseous hydrogen (H_2) **(1)**. ~~Suppose 68.5kg $CO_{(g)}$ is reacted with 8.60kg $H_{2(g)}$.~~ What would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$?

- **(1, 2, 4)** Knowing that methanol is made when you react CO with H_2, what would be the theoretical yield of methanol if you mix 68.5kg of $CO_{(g)}$ with 8.60 $H_{2(g)}$? ~~Calculate the theoretical yield of methanol.~~ ~~If 3.57 x 10^4g CH_3OH is actually produced,~~
- **(2, 4)** You found that that 3.57x10^4g CH_3OH is actually produced, what is the percent yield of methanol?

Content difficulty: Medium

Source:
Zumdahl, S.S. & Zumdahl, S.A. (2008) *Chemistry* (8th ed.). Belmont, CA: Brooks Cole Press. (p. 113).

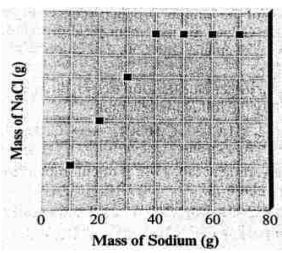
Question 2

~~You have seven closed containers, each with equal masses of chlorine gas (Cl_2).~~ There are seven total closed gas jars shown in the image above. Each jar has the same mass of chlorine gas

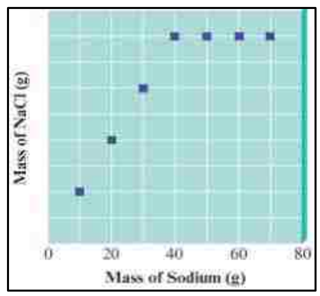
($Cl_{2(g)}$) in it **(1)**. ~~You add 10.0g of sodium to the first sample, 20.0g of sodium to the second sample, and so on.~~ You add sodium, ($Na_{(s)}$), as follows: **(1,4,5)**

Jar 1: 10.0g $Na_{(s)}$
Jar 2: 20.0g $Na_{(s)}$
Jar 3: 30.0g $Na_{(s)}$
Jar 4: 40.0g $Na_{(s)}$
Jar 5: 50.0g $Na_{(s)}$
Jar 6: 60.0g $Na_{(s)}$
Jar 7: 70.0g $Na_{(s)}$

Sodium and chlorine react to form sodium chloride. After the reaction in each jar is complete, you collect and measure the amount of sodium chloride formed. The graph below shows your data. **(10, 11)**



(Original graph)



(9)

Answer the following questions about this reaction and the data in the graph: **(5,11)**

a. Write the balanced equation for the reaction between Cl_2 gas and sodium.
b. ~~Calculate the mass of NaCl formed when 20.0g of sodium is used.~~
c. If you notice, there are no numbers on the y-axis of graph above. This means that you will have to figure out the numbers on the y-axis. Based on the balanced equation in part a, what mass of NaCl should be produced when 20.0g of Na reacts with the Cl_2 in jar 2? **(5,10,10)**
d. Based on the amount of NaCl that you calculated in part b, label the y-axis of your graph. **(7,10,11)**
e. ~~Calculate the mass of Cl_2 in each container.~~
f. Use the balanced equation from part a to calculate the mass of chlorine gas you would need to react with the 50g of Na in jar 5. **(7,10,11)**
g. Each of the jars has the same mass of Cl_2. We do not know what that mass is, but we do know that there was enough Cl_2 to react with the 20.0g of Na in jar 2. What mass of Cl_2 was needed to produce the amount of NaCl you calculated in step b of this problem?
h. ~~Identify the remaining reactant, and determine its mass for parts b and d above.~~
i. How does the amount of Cl_2 you calculated in part e compare to the amount of Cl_2 that is actually in the jar? What is the leftover reactant?
j. Calculate the mass of NaCl formed when 50.0g (jar 5) of sodium is used.
k. What is the mass of the leftover reactant in jar 5?
l. ~~Explain the shape of the graph.~~ **(5)**

m. In the graph above, you can see that more NaCl is produced when as you add more Na, but when you add more than 40g of Na, the amount of NaCl produced is does not change. Using the information you found in the previous steps, explain the following:
   a. Why the amount of NaCl produced increases at first when you add more sodium?
   b. Why is the amount of NaCl produced constant when you add more than 40g of sodium? **(1,2,10,11)**

Content difficulty: Hard

Source:
Zumdahl, S.S. & Zumdahl, S.A. (2012) *Chemistry: An atoms first approach* (2^{nd} ed.). Boston, MA: Cengage Learning. (p. 241).

Question 3

When ethane (C_2H_6) reacts with chlorine (Cl_2), the main product is C_2H_5Cl as the main product. **(1,3)** ~~but other products containing Cl, such as $C_2H_4Cl_2$, are also obtained in small quantities.~~ Other products containing chlorine are also produced in small amounts. ~~formation of these other products reduces the yield of C_2H_5Cl.~~ These are called minor products. Minor products decrease the amount of C_2H_5Cl (main product) that is made in this reaction **(1,3).** ~~Calculate the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2, assuming that C_2H_6 and Cl_2 react only to form C_2H_5Cl and HCl. Calculate the percent yield of C_2H_5Cl if the reaction produces 206 g C_2H_5Cl.~~

1. Write the balanced equation for the reaction between ethane and chlorine that produces C_2H_5Cl and HCl as products. **(10,11)**
2. Using the balanced equation you found in part 1, find the theoretical yield of C_2H_5Cl when 125g of C_2H_6 reacts with 255g of Cl_2? **(2,4,5,10)**
3. If conducted the reaction described in part 2 in the laboratory and only collected 206g of C_2H_5Cl, what is the percent yield of C_2H_5Cl? **(2,4,5, 10, 11)**

Content difficulty: Medium

Source:
Brown, T.L., LeMay, H. E., Bursten, B.E., Murphy, C.J., Woodward, P.M. & Stoltzfus, M.W. (2015). *Chemistry: The central science* (13 ed.). Upper Saddle River, NJ: Pearson (p. 118).

Question 4

When hydrogen sulfide gas is bubbled into a solution of sodium hydroxide, the reaction ~~forms~~ makes **(1)** sodium sulfide and water. You are told that this reaction gives you a 92.0% yield for sodium sulfide when you do the reaction in the laboratory. **(1,5,11)**

How many grams of sodium sulfide ~~are formed~~ will you make in the laboratory if you mix **(1)** 1.25g of hydrogen sulfide ~~is bubbled into a solution containing~~ **(3)** with 2.00g of sodium hydroxide, ~~assuming that the sodium sulfide is made in a 92.0% yield?~~ **(10)**

Content difficulty: Hard

Source:
Brown, T.L., LeMay, H. E., Bursten, B.E., Murphy, C.J., Woodward, P.M. & Stoltzfus, M.W. (2015). *Chemistry: The central science* (13 ed.). Upper Saddle River, NJ: Pearson (p. 118).

Question 5

~~The Haber process is the conversion of nitrogen and hydrogen at high pressure into ammonia as the following:~~ You work in a factory that manufactures ammonia gas.**(11)** The factory uses the Haber process to make ammonia. In the Haber process, nitrogen gas and hydrogen gas react together at a high pressure to create ammonia gas. This process is shown in the reaction below: **(11)**

$$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)}$$

~~If you must produce 700g of ammonia, what mass of nitrogen should you use in the reaction, assuming that the percent yield of this reaction is 70%?~~ You know that this process will give you a percent yield of 70% for ammonia gas.**(1)** Your job is to make 700g of ammonia.**(11)** What mass of nitrogen do you need to use so you can collect 700g of ammonia? **(5,7,11)**

Content difficulty: Hard

Adapted from:
Brown, T.L., LeMay, H. E., Bursten, B.E., Murphy, C.J., Woodward, P.M., & Stoltzfus, M.W. (2015). *Chemistry: The central science* (13 ed.). Upper Saddle River, NJ: Pearson (p. 338).

Question 6

~~The raw material used as a source of chromium and chromium compounds is a chromium-iron ore called chromite.~~ Chromite, a chromium-iron ore, is the source of chromium and chromium compounds. **(1)** By mixing chromite ($FeCr_2O_4$) with sodium carbonate, Na_2CO_3, you get sodium chromate, Na_2CrO_4. **(1,3)** ~~For example, sodium chromate, Na_2CrO_4, is made by roasting chromite with sodium carbonate, Na_2CO_3. A simplified version of the net reaction is:~~ This reaction is shown below in the balanced equation **(10):**

$$4FeCr_2O_4 + 8Na_2CO_3 + 7O_2 \rightarrow 8Na_2CrO_4 + 2Fe_2O_3 + 8CO_2$$

What is the percent yield if 1.2kg of Na_2CrO_4 ~~is produced from ore that contains~~ 1.0kg of $FeCr_2O_4$? What is the percent yield of Na_2CrO_4 if 1.2kg of Na_2CrO_4 is made from 1.0kg of $FeCr_2O_4$? **(1)**

Content difficulty: Easy

Source:
Bishop, M. (2001). An Introduction to Chemistry. Chiral Publishing Company.

Question 7

~~How many grams of calcium hydroxide are needed to make 3.75 grams of calcium phosphate?~~ Calcium hydroxide ($Ca(OH)_2$) and phosphoric acid (H_3PO_4) react to produce calcium phosphate ($Ca_3(PO_4)_2$) and water. **(4,10,11)**

A. Write the balanced equation for this reaction. **(10,11)**

B. How many grams of calcium hydroxide, ($Ca(OH)_2$), do we need to measure out to make 3.75g of calcium phosphate ($Ca_3(PO_4)_2$)? **(1,9,10)**

~~What if a mistake was made and we used 2.37 grams of $Ca(OH)_2$ and 2.69 grams of H_3PO_4? How much $Ca_3(PO_4)_2$ would be obtained?~~

C. We made a mistake in our measurements. Instead of using the amount of calcium hydroxide you calculated in Part B, we used 2.37g of $Ca(OH)_2$ and 2.69g of H_3PO_4. How many grams of $Ca_3(PO_4)_2$ should be made under these conditions? **(10,11)**

D. Which compound is the limiting reagent for the reaction described in Part C? **(10,11)** ~~What if 2.37 grams of $Ca(OH)_2$ were used along with excess H_3PO_4 and only 2.98 grams of $Ca_3(PO_4)_2$ were obtained (instead of the theoretical yield)? What would be the percent yield?~~

E. What would be the percent yield if we only collected 2.98g of calcium phosphate when we did the reaction described in Part C in a laboratory? **(1,3,7,10)**

Content difficulty: Easy

Source:
Instructor's test bank. Item received via internal correspondence from a general chemistry instructor.

APPENDIX D

Student Interview Guide

I. Informed Consent

a. Hi__________________, my name is Eshani, and I want to thank you for taking time to meet with me today. I appreciate your insights and help with this project. Before we begin, did you receive the Informed Consent document I emailed you?

i. Did you have a chance to read it?
ii. Do you have any questions about it?
iii. Do you agree to participate in this interview? Do you agree to be audio- and video-taped during the interview? If so, could you please sign this copy of the Informed Consent document? Thank you.

II. Background Questions

a. Can you tell me a little bit about your personal background?

i. Where are you from?
ii. Which language(s) did you grow up speaking?
iii. How old were you when you came to the U.S.?
iv. [If appropriate, omit for NES.] How much English did you speak before you moved here?
v. [If applicable] Tell me a little bit about your experiences moving to the U.S without speaking English.
[Add personal background here: “I also did not speak English growing up.

My family and I immigrated in to the U.S. when I was 12...”]

b. Tell me about your chemistry coursework.

i. Which chemistry course are you taking now? (Which other courses have you taken? Which ones do you still have to take?)
ii. What is your favorite topic in chemistry? What is your least favorite topic?
iii. How do you feel about the exams or quizzes you’ve been taking in your chemistry class?
iv. What do you find most difficult about chemistry test questions?
v. What do you think could be done to make the exams easier for you?

vi. As you probably know by now, I am a graduate student in the chemistry department, and I am interested in understanding how our ability to speak and read English affects the way we interpret test questions. Since English is not my first language, taking tests was particularly challenging for me. I am very interested in finding out what you think about certain test questions that I will show you today. Just so you know, you are not being graded in any capacity on these questions, and the goal is not to get the right answer. I am more interested in your thoughts: what you think about how the question is presented to you and how you would solve these questions. For example, you'll start by reading the question and talk me through the process of how you would solve the question. Then, you can tell me about any phrases or words that may be unclear or sentences that are difficult to understand, or anything about the question that is good or bad in terms of helping you solve it. The information I get from this interview will help me understand how to change chemistry exam questions so students them better.

III. Think aloud activity

a. I am going to show you some chemistry exam questions and ask what you think of them. I particularly want to know which parts of the questions make them easy or hard to comprehend. After I give you each question, I want you to read it out loud. I want to remind you that you are not being graded today on your answer choices. I am more interested in your process of solving the problems (like how you set it up and your approach to solving it) and what you think about the specific parts of the questions. So, shall we get started with the first question?

[Present first test question to the participant. *Do not mention if it is an original or modified version of the question*]

i. Please take a few minutes to read the question out loud.

ii. What is the problem asking you to do? Is it easy to figure out what this question is asking? Why or why not? Was the question easy? Was the question hard?

iii. Can you talk me through how you would set up and solve this problem? Feel free to write down your work.

1. I saw that you started your work with __________. Can you explain why you started here? (Note: Ask these follow up questions based on how the student set up the problem)

iv. Please circle any words or phrases that you found unclear with the red pen. Please circle any words that you found to be helpful with the blue pen. Here is the legend for your reference. (If applicable) Why did you circle ______ ?

1. [Possible follow up question when applicable: What do you think the question means by the phrase: ________ ?]
2. Where do you think someone else could get stuck on this problem?

Ok, now I am going to show you another version of the same question and ask you to think about how it compares to the first one you just saw.

[Take the first of the question version back. Present the next version of the same question to the participant. *Do not mention that it is an original or modified version of the question*]

v. Please take a minute to read through the question out loud.

vi. Please circle any words or phrases that you found unclear with the red pen. Please circle any words that you found to be helpful with the blue pen? (If applicable) Why did you circle ______ ?

1. [Possible follow up question when applicable: What do you think the question means by the phrase: ________ ?]
2. Where do you think someone else could get stuck on this problem?

[Give back the participant's copy of first version of the question.]

vii. Now that you have seen both versions of the same question, which version was easier for you to understand – the first or the second? Can you explain why?

viii. Which version would you prefer to have on the test? Can you tell me why?

ix. Neither one of the two question versions is perfect. If you could rewrite this question so that it would be easier for you to understand, how would you do it? Feel free to add, remove, or change anything in the question so the question is easier to understand.

x. [Look at features of the questions that have changed by the participant if applicable] Why did you change this word or phrase? Why did you add the following ______ ? Why did you remove the following _______ ?

xi. Thank you for talking through the first question. Now, I'll ask you the same set of questions as we go through the next set of questions.

[Repeat questions from Parts a(i) – a(xi) for the additional sets of test item pairs, as time allows.]

IV. Post-activity questions

a. Now that we've gone through all the problems, I would like to ask you a few general questions about the problems you just completed.

    i. Of the three sets of questions that you saw here today, which one was the most difficult to understand? Why?
    ii. Which question was the most clear in what was being asked of you to solve the problem? Why?
    iii. What advice would you give to your professor about how they could write chemistry exam questions so that you can understand them better?
    iv. Do you have any questions? Any comments or additional thoughts about what we have talked about here?

Thank you very much for your time and willingness to participate in this project.

Pen Colors Legend

- Unclear = Red
- Helpful = Blue

APPENDIX E

- 1 - Generated on IRBNet
UNLV Social/Behavioral IRB - Exempt Review

Exempt Notice

DATE: June 7, 2017
TO: MaryKay Orgill
FROM: Office of Research Integrity - Human Subjects

PROTOCOL TITLE: [1069437-1] The Equitable Assessment of English Language Learners in General Chemistry

ACTION: DETERMINATION OF EXEMPT STATUS
EXEMPT DATE: June 7, 2017
REVIEW CATEGORY: Exemption category #2

Thank you for your submission of New Project materials for this protocol. This memorandum is notification
that the protocol referenced above has been reviewed as indicated in Federal regulatory statutes 45CFR46.101(b) and deemed exempt.

We will retain a copy of this correspondence with our records.

PLEASE NOTE:

Upon final determination of exempt status, the research team is responsible for conducting the research as stated in the exempt application reviewed by the ORI - HS and/or the IRB which shall include using the most recently submitted Informed Consent/Assent Forms (Information Sheet) and recruitment materials.

If your project involves paying research participants, it is recommended to contact Carisa Shaffer, ORI Program Coordinator at (702) 895-2794 to ensure compliance with the Policy for Incentives for Human Research Subjects.

Any changes to the application may cause this protocol to require a different level of IRB review. Should any changes need to be made, please submit a **Modification Form**. When the above-referenced protocol has been completed, please submit a **Continuing Review/Progress Completion report** to notify ORI -HS of its closure.

If you have questions, please contact the Office of Research Integrity - Human Subjects at IRB@unlv.edu or call 702-895-2794. Please include your protocol title and IRBNet ID in all correspondence.

Office of Research Integrity - Human Subjects
4505 Maryland Parkway . Box 451047 . Las Vegas, Nevada 89154-1047
(702) 895-2794 . FAX: (702) 895-0805 . IRB@unlv.edu

Informed consent

Department of Chemistry and Biochemistry

TITLE OF STUDY: The equitable assessment of English language learners in general chemistry.

INVESTIGATOR(S): Dr. MaryKay Orgill (UNLV Professor), and Eshani Lee (UNLV Doctoral Student)

For questions or concerns about the study, you may contact Dr. Orgill at **702 895-3580.**

For questions regarding the rights of research subjects, any complaints or comments regarding the manner in which the study is being conducted, contact **the UNLV Office of Research Integrity – Human Subjects at 702-895-2794, toll free at 877-895-2794 or via email at IRB@unlv.edu.**

Purpose of the Study

You are invited to participate in a research study. The purpose of this study is to determine what undergraduate general and organic chemistry students understand about general chemistry assessment questions.

Participants

You are being asked to participate in the study because you fit this criterion: you are currently enrolled in the first semester of general chemistry and/or you are a student whose first language is not English.

Procedures

If you volunteer to participate in this study, you will be asked to do the following: participate in a one-hour interview about how you understand question exams on theoretical yield and limiting reagents.

Benefits of Participation

There may not be direct benefits to you as a participant in this study. However, we hope to learn about how students' interpret general chemistry test items, and discussing your thoughts on this topic may change your understanding of it.

Risks of Participation

There are risks involved in all research studies. This study may include only minimal risks which may include embarrassment, emotional distress, and psychological trauma associated with the discussion of unfamiliar topics.

Cost /Compensation

There will not be financial cost to you to participate in this study. The study will take approximately 60 minutes of your time. You will be provided with a $10 Amazon.com gift at the end of the interview.

Confidentiality

All information gathered in this study will be kept as confidential as possible. No reference will be made in written or oral materials that could link you to this study. All records will be stored in a locked facility at UNLV for three years after completion of the study. After the storage time the information gathered will be destroyed.

Voluntary Participation

Your participation in this study is voluntary. You may refuse to participate in this study or in any part of this study. You may withdraw at any time without prejudice to your relations with

UNLV. You are encouraged to ask questions about this study at the beginning or any time during the research study.

Participant Consent:

I have read the above information and agree to participate in this study. I have been able to ask questions about the research study. I am at least 18 years of age. A copy of this form has been given to me.

____ ___________________

Signature of Participant Date

____

Participant Name (Please Print)

Audio/Video Taping:

I agree to be audio and video taped for the purpose of this research study.

____ ___________________

Signature of Participant Date

____

Participant Name (Please Print)

APPENDIX F

In-Class Solicitation Script

Good (morning/afternoon),

I would like to thank Professor ___________________ for letting me come in and speak with you today.

My name is Eshani Lee, and I am a doctoral student in the chemistry department here at UNLV. I am working on my dissertation research and am looking for general chemistry students to participate in a one-hour audiotaped individual interview. You will receive a $10 gift card for Amazon.com for your participation immediately following the interview.

The interview will be held in the chemistry building. I am willing to work with your schedules and find a time that best works for you.

If you agree to be interviewed, the interview questions I will ask you do not have right or wrong answers, rather they will ask you for your perspectives about how you understand general chemistry test questions. Through this research I am hoping to find ways to improve the way exam questions are written in undergraduate general chemistry courses. To accomplish this goal, I need to hear from general chemistry undergraduate students about their experiences with these exam questions, specifically how you read and interpret chemistry test questions.

My research is comparative in nature, so I am looking for general chemistry undergraduate students who fit two different study criteria:

First, I am looking for students who grew up speaking English as a second language AND who have been in the U.S. for less than 8 years.

Second, I am looking for all other students (i.e. students who are native English speakers from anywhere in the world including the U.S.).

> If you are interested in participating please write your name and preferred e-mail address on the index card I handed out. If you do not want to participate, leave the card blank. Pass back all cards. If I get a card with your contact information on it, I will contact you to set up a specific day/time for us to meet for the interview.

Thank you all for your time, and a special thank you to anyone who chooses to participate.

REFERENCES

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment, 8*(3), 231-257.

Abedi J. (2007). English language proficiency assessment and accountability under NCLB Title III: An overview. In *English Language Proficiency Assessment in the Nation: Current Status and Future Practice* (chapter 1). Retrieved from http://education.ucdavis.edu/sites/main/files/ELP_Report.pdf.

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification* (CSE Report 666). National Center for Research on Evaluation, Standards, and Study Testing (CRESST). Los Angeles, CA: University of California, Los Angeles.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16-26.

Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Abedi, J., & Lord, C. (2011). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219-234.

Akkuzu, M., & Uyulgan, M.A. (2017). Step by step learning using the I diagram in the systematic qualitative analysis of cations within a guided inquiry learning approach. *Chemistry Education Research and Practice, 18*, 641-658.

American Association for the Advancement of Science (AAAS). (1989). *Science for all Americans.* New York: Oxford University Press.

American Institute for Research (2012). *Supporting English Language Learners: A pocket guide for State and District leaders.* Washington, DC. Available at http://www.air.org/sites/default/files/downloads/report/ELL_Pocket_Guide1_0.pdf.

Amos, Z.S. (2009). *The relationship of readability on the science achievement test: A study of 5th grade achievement performance* (master's thesis). Bowling Green State University, Bowling Green, Ohio.

Astin, A. W., & Astin, H.S. (1992). *Undergraduate science education: The impact of different college enrviornments on the educational pipeline in the sciences*. Los Angeles, CA: University of California, Graduate School of Education, Higher Education Research Institute.

Bajak, A. (2014). Lectures aren't just boring, they are ineffective, too, study finds. *Science Magazine*. Retrieved from http://www.sciencemag.org/news/2014/05/lectures-arent-just-boring-theyre-ineffective-too-study-finds

Boonen, A.J.H., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research, 68*, 15-26.

Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Teaching and Learning in Higher Education: New Trends and Innovations,* 1-9.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20*, 463–494.

Bodner, G., & Orgill, M. (2007). *Theoretical frameworks for research in chemistry/science education*. Prentice Hall.

Burke, R.J., & Mattis, M.C. (Eds.). (2007). Women and minorities in Science, Technology, Engineering and Mathematics. Northampton, MA: Edward Elgar.

Brown, C. L. (2005). Equity of Literacy-Based Math Performance Assessments for English Language Learners. *Bilingual Research Journal, 29*(2), 337.

Bruna, K.R., Vann, R., & Escudero, M.P. (2007). What's language got to do with it? A case study of academic language instruction in a high school "English learner science" class. *Journal of English for Academic Purposes*, *6*, 36-54.

Carrell, P. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19(4), 727-752.

Carter C.S., & Brickhouse, N.W. (1989). What makes chemistry difficult? Alternate perceptions. *Journal of Chemical Education*, *66*(3), 223-225.

Change the Equation Vital Signs: Reports on the condition of STEM learning in the U.S. (2010) *STEM Help Wanted: Demand for Science, Technology, Engineering and Mathematics weathers the storm*. Retrieved August 30, 2016, from http://changetheequation.org/sites/default/files/CTEq_VitalSigns_Supply%20(2).pdf

Chatmot, A.U., & O'Malley, J.M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach.* New York: Longman.

Collier, L. (2008). The importance of academic language for English language learners. *The Council Chronicle: National Council of Teachers of English*.

Collier, V.P (1995). Acquiring a second language for school. *Directions in Language & Education,* National Clearinghouse for Bilingual Education, *1*(4), 1-8.

Center for Research on Evaluation & Student Testing [CRESST]. (2001). *Policy brief no. 4.* Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Creswell, J.W. (2007). *Qualitative inquiry and research design: Choose among give approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 405–438.

Cummins, J. (1980). Psychological assessment of immigrant children: Logic or intuition? *Journal of Multilingual and Multicultural Development*, *1*, 97-lll.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, *11*, 132–149.

Cummins, J. (1984) *Bilingual Education and Special Education: Issues in Assessment and Pedagogy* San Diego: College Hill

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.

Denzin, N. K. (1978). *The research act* (2nd ed). New York: McGraw-Hill.

de Schonewise, E. A., & Klingner, J. K. (2012). Linguistic and Cultural Issues in Developing Disciplinary Literacy for Adolescent English Language Learners. *Topics in Language Disorders, 32*(1), 51-68.

Downing, S. M. (2002). Construct-irrelevant variance and flawed test questions: Do multiple-choice item writing principles make any difference? *Academic Medicine, 77*(10), S103-104.

Education Commission of the States. (2014). *How is an "English language learner" defined in state policy? 50-State Comparison.* Denver, CO: ECS.org. Available at http://ecs.force.com/mbdata/mbquestNB2?rep=ELL1402

Fitzgerald, J. (1995). English-as-second-language learners' cognitive reading processes: A review of research in the United States. *Review of Educational Research*, *65*, 145-190.

Fong, S.R., & Siegel, M.A. (2005). Teaching well: Science teachers' investigation and use of student sociocultural background. In D.M. McInerney & S. van Etten (Eds.), *Focus on teaching* (101-128). Greenwich, CT: Information Age Publishing.

Garcia, G., & Pearson, P. (1994). Assessment and diversity. *Review of Research in Education, 20*, 337–391.

Goubeaud, K. (2010). How is science learning assessment at the postsecondary level? Assessment and grading practices in college biology, chemistry and physics. *Journal of Science Educational Technology*, *19*, 237-245.

Haladyna, T. M. D., S.M. (2004). Construct-Irrelevant Variance in High-Stakes Testing. *Educational Measurement: Issues and Practice*, 17-27.

Haraway, D. (1989). *Primate visions*. New York: Routledge.

Haraway, D. (1991). *Simians, cyborgs, and women*. New York: Routledge.

Haraway, D. (1999). *Modest witness @ second millennium*. New York: Routledge.

Hartman, J.R., Lin, S. (2011). Analysis of student performance on multiple-choice questions in general chemistry. *Journal of Chemical Education*, *88*, 1223-1230.

Ji, L., Zhang, Z., & Nisbett, R.E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology, 87*(1), 57-65.

Kanno, Y., & Cromley, J. (2013). English language learners’ access to and attainment in Postsecondary Education. *TESOL Quarterly, 47*(1), 89-121.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English Language Learners Taking Large-Scale Assessments: A Meta-Analysis on Effectiveness and Validity. *Review of Educational Research*, *79*(3), 1168-1201.

Kiray, G. (2015). Macro-structure analysis of reading comprehension paragraphs of KPDS and YDS exams within years 2003-2013. *Hasan Ali Yücel Eğitim Fakültesi Dergisi Cilt*, *23*(12-1), 219-233.0.

Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, DC: Council of Chief State Officers.

Lacroix, N. (1999). Macrostructure construction and organization in the processing of multiple text passages. *Instructional Science*, *27*, 221-233.

Lo, L., HO, C.S., Wong, Y.K., Chan, D.W., Chungg, K.K. (2016). Understanding the microstructure and the macrostructure of passages among Chinese elementary school children. *Journal of Psycholinguistic Research*, *45*(6), 1287-1300.

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

LeClair, C., Doll, B., Osborn, A., Jones, K. (2009). English language learners’ and non-English language learners’ perceptions of the classroom environment. *Psychology in the Schools*, *46*(6), 568-577.

Lee, O. (2001). Culture and language in science education: What do we know and what do we need to know? *Journal of Research in Science Teaching*, *38*(5), 499-501.

Lee, O., and Fradd, S. (1998). Science for All, including students from non-English-language backgrounds. *Educational Researcher*, *27*(4), 12-21.

Lee, O. (2004). Teacher change in beliefs and practices in science and literacy instruction with English language learners. *Journal of Research in Science Teaching, 41*, 65–93.

Lemke, J.L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.

Lemke, J.L. (2001). Articulating communities: Sociocultural perspectives on science education. *Journal of Research in Science Teaching, 38*(3), 296-316.

Lewis, S.E., Shaw, J.L., & Freeman, K.A. (2010). Creative exercises in general chemistry: A student-centered assessment. *Journal of College Science Teaching, 40*(1), 48-53.

Leontev, A. N. (1978). *Activity, Consciousness, and Personality*. Prentice-Hall. (Trans. M. J. Hall).

Lincoln, Y.S., & Guba, E.G. (1985). Establishing trustworthiness. In Y.S. Lincoln & E.G. Guba (Eds.), *Naturalistic Inquiry* (pp. 289-331). Newbury Park, CA: Sage.

Luykx, A., Lee, O., Hart, J. (2007). Cultural and home language influences on children's responses to science assessments. *Teachers College Record*, *109*(4), 897-926.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in Math tests. *Educational Assessments*, *14*, 160-179.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). Phoenix, AZ: American Council on Higher Education and Oryx Press

Michaels, S., & O'Connor, M. C. (1990). *Literacy as reasoning within multiple discourses: Implications for policy and educational reform.* Paper presented at the Council of Chief

State School Officers Summer Institute on Restructuring Learning, Education Development Center, Literacies Institute, Newton, MA.

Miller, E. R., Okum, I., Sinai, R., & Miller, K. S. (1999, April). *A study of the English language readiness of limited English proficient students to participate in New Jersey's statewide assessment system.* Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal.

Mokhtar, C. (2012). Clark County School District's English Language Learners: An analysis of enrollment, educational opportunities, and outcomes in Nevada and CCSD. Provided as Internal Working Document to The Lincy Institute at the University of Nevada, Las Vegas by the Annenberg Institute for School Reform at Brown University

Museus, S.D., Palmer, R.T., Davis, R.J., & Maramba, D.C. (2011). Racial and ethnic minority students' success in STEM education. *American Society of Higher Education Report, 36*(6), 1-139.

National Center for Education Statistics, (2004). *English language learner students in U.S. public schools: 1994 and 2000*. Washington, DC: U.S. Department of Education Institute of Education. Available at https://nces.ed.gov/pubs2004/2004035.pdf.

National Center for Education Statistics (2012). *The Nation's Report Card: Reading 2011*. Washington, DC: U.S. Department of Education Institute of Education Sciences. Nevada State Senate. 2011. Nevada Education Data Book. Available at http://www.leg.state.nv.us/Division/Research/Publications/EdDataBook/2011/Ch01.pdf .

National Science Teachers Association [NSTA]. (2014). *Next generation science standards. Science and engineering practices.* Arlington, VA: NSTA.

Noble, T., Suarez, C., Rosebery, A., O'Conner, M.C., Warren, B., Hudicourt-Barnes. J. (2012). "I never thought of it as freezing": How students answer questions on large-scale science test and what they knew about science. *Journal of Research in Science Teacher, 49*(6), 778-803.

Pappamihiel, N., & Mihai, F. (2006). Assessing English language learners' content knowledge in middle school classrooms. *Middle School Journal, 38*(1), 34-43.

Peim, N. (2009). Activity theory and ontology. *Educational Review, 61*(2), 167-180.

Penuel, W. R., & Wertsch, J. V. (1995). Vygotsky and identity formation: A sociocultural approach. *Educational Psychologist, 30*, 83-92.

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife*, *6*, e27725. http://doi.org/10.7554/eLife.27725

President's Council of Advisors on Science and Technology [PCAST]. (2012). Engage to excel: Producing one million additional college graduate with degrees in Science, Technology, Engineering and Mathematics. Retrieved November 16, 2013 from http://www.whitehouse.gov/.

Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review, 29*, 892-900.

Readence, J., Bean, T., & Baldwin, R. S. (2004). Content area literacy: An integrated approach (8th ed.). Dubuque, IA: Kendall/Hunt Publishing.

Roth, W., & Lee, Y. (2007). "Vygotsky's neglected legacy:" Cultural-historical activity theory. *Review of Educational Research, 77*(2), 186-232.

Ryu, M. (2015). Positionings of racial, ethnic, and linguistic minority students in high school biology class: Implications for science education in diverse classrooms. *Journal of Research in Science Teaching, 52*(3), 347-370.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump*. Princeton, NJ: Princeton University Press.

Seymour, E., & Hewitt, N. (1994). Talking about leaving: Factors contributing to high attrition rates among science, mathematics, and engineering undergraduate majors. Boulder, CO: Bureau of Sociological Research, University of Colorado.

Siegel, M. A. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching, 44*(6), 864-881. doi:10.1002/tea.20176

Siegel, M. A., Menon, D., Sinha, S., Promyod, N., Wissehr, C., & Halverson, K. L. (2014). Equitable Written Assessments for English Language Learners: How Scaffolding Helps. *Journal of Science Teacher Education, 25*(6), 681-708. doi:10.1007/s10972-014-9392-1

Siegel, M.A., Wissehr, C. & Halverson, K. (2008). Sounds likes success: A framework for equitable assessment. *The Science Teacher*, 43-46.

Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices: An introduction. Basterra, In M.R., Trumbull, E., & Solono-Flores, G. (Eds.) Cultural validity in assessment: A guide for educators (pp. 3-21). New York: Routledge.

Solano-Flores, G., & Li, M. (2008). Examining the Dependability of Academic Achievement Measures for English Language Learners. *Assessment for Effective Intervention, 33*(3), 135-144.

Solano-Flores, G., & Li, M. (2009). Language Variation and Score Variation in the Testing of English Language Learners, Native Spanish Speakers. *Educational Assessment*, *14*(3-4), 180-194.

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

Solano-Flores, G., & Trumbull, E. (2003). Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners.

Solorzano, R. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, *78*(2), 260-329.

Snow, C.E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328*, 450-452.

Teachers of English to Speakers of Other Languages. (2008). *TESOL/NCATE Standards for the recognition of initial programs in P-12 ESL teacher education*. Alexandria, VA: Author. Retrieved June 2, 2009, from http://www.tesol.org/s_tesol/sec_document.asp?CID=219&DID=10698.

Thompson, S., & Thurlow, M. (2002). *Universally Designed Assessments: Better tests for everyone! NCEO Policy Directions.* National Center on Educational Outcomes.

Tobias, S. (1990). *They're not dumb, they're different: Stalking the second tier*. Tucson, AZ: Research Corporation.

Trumbull, E. & Solano-Flores, G. (2011). The role of language in assessment. In M. Rosario Basterra, E. Trumbull & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (22-45). New York, NY: Routledge.

Walqui, A. (2003). *Conceptual framework: Scaffolding instruction for English learners*. San Francisco, CA: Cambridge University Press.

White, R.T., & Gunstone, R. (1992). *Probing understanding*. London, UK: Routledge.

Woldeamanuel, M.M., Atagana, H., & Engida, T. (2014). What makes chemistry difficult? *African Journal of Chemical Education*, *4*(2), 31-43.

Wygoda, L., Teague, R. (1995). Performance-based chemistry: Developing assessment strategies in high school chemistry. *The Trading Post, 72*(10), 909-911.

CURRICULUM VITAE

ESHANI N. LEE

Professional Address:

Department of Chemistry & Biochemistry
University of Nevada, Las Vegas
4505 S. Maryland Pkwy., Box 4003
Las Vegas, NV 89154-4003

Contact:
EGandhi12@gmail.com

Educational Background

Ph.D. Chemistry (Chemistry Education) May 2018
University of Nevada, Las Vegas, Nevada
Advisor: Dr. MaryKay Orgill
Dissertation Topic: *Breaking the Language Barrier: Equitable Assessment in General Chemistry*
Defense Date: May 7th, 2018

Mentorship Certification May 2018
Graduate College, University of Nevada, Las Vegas, Nevada
Professional Development: *Skills & Knowledge to Mentor in Higher Education*

M.A.S. Arts & Science (Biology) August 2012
University of Nevada, Las Vegas, Nevada
Thesis Topic: *Biochemical Adaptations in Fish*

B.A. Social Sciences June 2005
University of California Irvine, California

Professional Experience

Research Experience

Graduate Researcher in Chemistry Education 2012 – 2018
Department of Chemistry & Biochemistry, University of Nevada, Las Vegas

Completed relevant doctorate level coursework in the fields of chemistry and education. Successfully passed comprehensive exams (oral & written) and defended prospectus. Conducted a quantitative preliminary study to examine the perceptions of English language learner (ELL) students on their learning experiences in a general chemistry course. Developed an in-depth dissertation research project that focused on how ELL and Native English Speaking undergraduate students interpret commonly used exam questions in general chemistry. Designed modified exam questions based on an assessment framework for equity. Conducted the analyses of qualitative and quantitative data using established analytical frameworks. Presented findings at conferences.

Graduate Research Assistant 2012-2016
GEAR UP UNLV STEM Grant
Gaining Early Awareness & Readiness for Undergraduate Programs
Office of the Provost, University of Nevada, Las Vegas

Involved in the implementation of professional development courses for Nevada's middle school in-service STEM teachers. Actively involved in designing STEM learning experiences that are applicable for teachers in their own classrooms. Worked alongside with discipline experts from life sciences, chemistry, engineering and mathematics to deliver STEM teaching lesson plans and educational technology tools that are integrative and engaging for students. As part of this program, collected qualitative data about faculty perceptions on recruitment and retention issues in STEM fields as well as faculty perceptions on student success and persistence in STEM fields. Conducted campus-wide interviews, transcribed, analyzed data, published in peer-reviewed STEM education journals and presented findings at conferences.

Curriculum Development for General Chemistry Laboratories 2016
Department of Chemistry & Biochemistry, University of Nevada, Las Vegas

Lead the redesigning and implementation of new general chemistry laboratory curriculum for over 300 undergraduate chemistry majors. Involved in the selection of research-based experiments and obtaining undergraduate student feedback on concepts, content, and procedures. Manage the generation of costs and waste for each new laboratory protocol implementing, which involved working closely with the stockroom and teaching assistants.

Graduate Researcher in Biology 2009-2012
School of Life Science, University of Nevada, Las Vegas

Conducted physiological research in the field of biochemical adaptations and managed endangered marine pupfish populations in laboratory. Performed routine animal care and husbandry protocols. Analyzed physiological responses of pupfish during developmental periods to temperature fluctuations using Strathkelvin instruments. Completed Masters' thesis.

Clinical Research Assistant 2003-2005

University of California, Irvine Medical Center

Performed structured interviews with selected participant groups to diagnose with psychiatric disorders in order to study perceptual anomalies. Evaluated sensory gating brain mechanisms using electrographic brain wave instruments.

Teaching Experience

Instructor 2017-present
Department of Chemistry & Biochemistry, University of Nevada, Las Vegas

Instructor of record for a 3-credit preparatory chemistry course for science majors. Designed a virtual, interactive course experience using online chemistry learning platforms, with added multimedia support to understand foundational topics in general chemistry. Developed online homework assignments, and multi-format course assessments. Implemented weekly group activities designed to encourage collaboration and discussion of scientific concepts.

Adjunct Faculty 2016-present
School of Liberal Arts & Sciences, Nevada State College, Las Vegas

Instructor for an online 3-credit introductory chemistry course for non-majors: Chemistry, Man & Society. Designed a virtual, interactive course experience using Mastering Chemistry through the Canvas learning management system. Developed online homework assignments, multimedia tutorials, and multi-format assessments. Integrated online class discussions on relevant issues and topics in chemistry that impact society and our everyday lives. Introduced virtual multimedia content support, which included video lessons, mini-video clips on how to solve chemical problems and short lectures.

Adjunct Faculty 2016-2017
School of Nursing, Arizona College, Las Vegas

Instructor for an accelerated, hybrid 4-credit Introductory General, Organic and Biochemistry undergraduate course for pre-nursing students with laboratory. Responsible for adapting the course design and curriculum to match current content standards. Developed online tutorials and discussions using Lecture Capture on the Blackboard learning management system. Created lessons using research-based instructional approaches that make chemistry more accessible to underrepresented minority students. Developed student-centered lectures and collaborative learning activities during class. Created all assessments, implemented laboratory lessons and discussions regarding current issues in chemistry. Held office hours and facilitated additional instructional support groups.

Adjunct Faculty 2014-present

School of Liberal Arts & Science, Nevada State College, Las Vegas

Instructor for a 4-credit Introductory General, Organic and Biochemistry undergraduate course with laboratory. Responsible for designing the course for pre-nursing students on the Canvas learning management system. Created lessons using research-based instructional approaches that make chemistry more accessible to underrepresented minority students. Developed student-centered lectures and collaborative learning activities during class. Created all assessments and implemented laboratory lessons. Held office hours and facilitated additional instructional support groups.

Recitation Lecturer 2014-present
Department of Chemistry & Biochemistry, University of Nevada, Las Vegas

Taught interactive lessons to undergraduate nursing students enrolled in the introductory General, Organic and Biochemistry course during weekly recitation sessions. Designed lectures for a diverse group in a class size of 260+ that focused on collaborative learning and team discussions. Designed and conducted team-based learning lessons, designed assessments, reviewed and graded course exams. Held weekly office hours.

Teaching Assistant 2012 – 2013
Department of Chemistry & Biochemistry, University of Nevada, Las Vegas

Taught the laboratory component for the general chemistry course for majors and non-majors. Demonstrated concepts and techniques for each lab exercise. Collected and evaluated weekly lab reports. Responsible for student safety practices and the management of laboratory equipment. Set up and graded lab mid-term exams and practicums in the WebCampus learning management system. Held weekly office hours.

Teaching Assistant 2010 – 2012
School of Life Sciences, University of Nevada, Las Vegas

Taught the laboratory component of the fundamentals of biology course for biology majors. Presented and demonstrated proper biological research techniques and broader biological concepts. Responsible for student safety practices and the management of laboratory equipment. Designed weekly quizzes and graded exams in the WebCampus learning management system. Held weekly office hours.

Teaching Assistant 2010
School of Life Sciences, University of Nevada, Las Vegas

Instructed students in the anatomy and physiology lab, and graded lab reports and quizzes. Taught methods and protocols for physiological tools, demonstrated dissections of animal tissues. Responsible for student safety practices and the management of laboratory equipment. Setup practicums and weekly quiz stations. Held weekly office hours.

Teaching Assistant 2009 – 2010
School of Life Sciences, University of Nevada Las Vegas.

Taught the laboratory component of a general biology class for non-majors. Prepared weekly assessments, graded exams and practicums. Demonstrated introductory biological concepts and performed as well as relevant applications. Responsible for student safety practices and the management of laboratory equipment. Held weekly office hours.

Publications

- **Gandhi-Lee, E.**, Skaza, H., Marti, E., Schrader, PG., Orgill, M. (2017). Faculty perceptions of recruitment and retention in STEM fields. *European Journal of STEM Education, 2:1.*
- **Gandhi-Lee, E.**, Skaza, H., Marti, E., Schrader, PG., Orgill, M. (2015). Faculty perceptions of the factors influencing success in STEM fields. *Journal of Research in STEM Education, 1(1), pp. 24-53.*

Grant/Funding Activities

Funded Activities

Graduate & Professional Student Association Travel Grant 2015
University of Nevada, Las Vegas, Las Vegas, NV.

Graduate & Professional Student Association Research Grant 2013
University of Nevada, Las Vegas, Las Vegas, NV.

Graduate & Professional Student Association Travel Grant 2012
University of Nevada, Las Vegas, Las Vegas, NV.

Honors, Awards & Recognition

The Graduate College Medallion Recipient *2018*
Medallion program honors exceptionally involved and high-achieving students.
University of Nevada, Las Vegas, NV

Featured in the UNLV Innovations Magazine:
Breaking the Language Barrier. (2017, May). *UNLV Newsletter @ UNLV Today*. Retrieved from https://www.unlv.edu/news/unlvtoday.

The Jean Nidetch Women's Center Scholarship *2017*
University of Nevada, Las Vegas, NV.

The Patricia Sastaunik Scholarship *2017*
University of Nevada, Las Vegas, NV.

The President's UNLV Foundation Graduate Research Fellowship *2016*
University of Nevada, Las Vegas, NV.

The Donna Weistrop and David B. Schaffer Scholarship *2015*
University of Nevada, Las Vegas, NV.

The Patricia Sastaunik Scholarship *2015*
University of Nevada, Las Vegas, NV.

Honorable Mention for Graduate Research Forum Presentation *2014*
University of Nevada, Las Vegas, NV.

The Board of Trustees and the Barrick Fellowships (Finalist) *2014*
University of Nevada, Las Vegas, NV.

Skills

- Graduate Student Mentorship Certificate
- Laboratory Hygiene and Safety
- Active Shooter Training
- CPR Certified
- Multilingual
- Strong written and oral communication skills
- Proficient in word processing, spreadsheets, database management, and Canvas, WebCampus and Blackboard learning management systems

Professional Societies & Memberships

- American Chemical Society (ACS)
- *National Association of Research in Science Teaching* (NARST)
- *National Science Teacher Association* (NSTA)
- *GPSA (Graduate & Professional Student Association)*
- *BIOS Graduate Association*

Conferences Attended

- 2017 Chemistry Education Research Graduate Student & Professional Development Conference, Miami, OH, June 2017.
- 2015 American Psychological Association Annual Convention, Toronto, Ontario, Canada, August 2015.
- 2014 Biennial Conference on Chemistry Education, Grand Rapids, MI, August 2014.
- 245^{th} National Meeting of the American Chemical Society, New Orleans, LA, April 2013
- 2013 Undergraduate Research Symposium, Festival of Communities, Las Vegas, NV, April 2013
- 2013 Graduate & Professional Student Association Research Forum, Las Vegas, NV, March 2013

Presentations

National/International

- **Gandhi-Lee, E**. & Orgill, M. (2017, June). *Equitable assessments of English language learners (ELLs) in general chemistry.* Presented at the Chemistry Education Research Graduate Student & Professional Development Conference, Miami, OH.
- Nealy, S., Carroll, K., Skaza, H., Marti, E., **Gandhi, E.**, Dulger, M., Gerrity, D., Olson, T., Schrader, P. G., & Orgill, M. (2015, November). *What Would an Alien Eat? Findings from an Inquiry Based Teacher Training.* Poster presented at the 24th NSF EPSCoR National Conference, Portsmouth, NH
- Wood, S., Orgill, M., Brennan, D., & **Gandhi, E.** (2012, August). *How do students' understandings of acids and bases change as they progress from general to organic chemistry?: A preliminary report.* Presented at the 22^{nd} Biennial Conference on Chemical Education, University Park, PA.
- **Gandhi, E**., Orgill, M., Schrader, P. (2013, April). *Nevada's GEAR UP: Developing and formalizing a needs analysis for professional development in STEM education.* Presented at the 245^{th} American Chemical Society, New Orleans, LA.
- Marti, E., Orgill, M., Schrader, P.G., **Gandhi, E.,** Curry, C., & Greene, F. (2013, August). *UNLV GEAR UP activities year one: Addressing STEM education in Nevada.* Poster presented at the 3rd Annual Colloquium on P-12 STEM Education: Research to Practice, University of Minnesota, St. Paul, MN.
- Nealy, S., Carroll, K., Skaza, H., **Gandhi, E**., Dulger, M., Gerrity, D., Olson, T., Nobles, D., Schrader, P., Orgill, M. (2014 August). *Design, development, and delivery of the Nevada GEAR UP STEM Summer Institute*. Presented at the 2014 Biennial Conference on Chemistry Education, Allendale, MI.
- **Gandhi, E.,** Skaza, H., Marti, E., Schrader, P., Orgill, M. (2014 August). *Faculty perceptions of the factors influencing success in STEM fields.* Presented at the 2014 Biennial Conference on Chemistry Education, Allendale, MI.
- Schrader, P.G., Orgill, M. K., Skaza, H., **Gandhi, E.,** Nealy, S., Marti, E., Dulger, M., Gerrity, D., Olson, T. A., Carroll, K., & Nobles, D. (2015, April). *What would an alien eat? Planning and implementing an intensive summer institute.* Paper presented to the annual meeting of AERA, April 16th-20th, Chicago, IL.

- **Gandhi-Lee, E.**, Kardash, C., Marchand, G., Bendixen, L., Nardi, N. (2015, August). *Literacy perceptions of international students in an undergraduate chemistry class.* Paper presented to the annual convention of APA, August 6th-9th, Toronto, Ontario, Canada.

Local/Regional

- **Gandhi, E.,** Orgill, M., Schrader, P., Marti, E. (2013, April). *GEAR UP: Addressing STEM Education in Nevada*. Presented at the 2013 Festival of Communities, Las Vegas, NV.
- **Gandhi, E.,** Orgill, M., Schrader, P. (2013, April). *Nevada's GEAR UP: Developing and formalizing a needs analysis for professional development in STEM education. Presented at the* 2013 Graduate & Professional Student Association Research Forum, Las Vegas, NV.
- Nealy, S., Carroll, K., Skaza, H., Marti, E., **Gandhi, E.,** Dulger, M., Gerrity, D., Olson, T., Schrader, P.G., & Orgill, M. (2014, August). Design, development, and delivery of the Nevada GEAR UP STEM Summer Institute. Poster presented at the American Chemical Society Southern Nevada Local Section 3rd Annual Research Poster Exhibition-Competition, Las Vegas, NV.
- Marti, E., Orgill, M., Schrader, P., **Gandhi, E.** (2014, March). *UNLV GEAR UP Activities Year One: Addressing STEM Education in Nevada*. Presented at the 2014 Graduate & Professional Student Research Forum, Las Vegas, NV.

Service

Committee Work

- Curriculum Development Committee (Department of Chemistry & Biochemistry). University of Nevada, Las Vegas, 2015-present.
- Chair, Constitution & Bylaws Committee (Graduate & Professional Student Association). University of Nevada, Las Vegas, 2014-present.
- Chair, Government Relations Committee (Graduate & Professional Student Association). University of Nevada, Las Vegas, 2013-present.
- Special Fees Committee (University-wide committee to review fee proposals). University of Nevada, Las Vegas, 2013-present.
- Executive Board (Graduate & Professional Student Association). University of Nevada, Las Vegas, 2013-present.
- Parking & Transportation Committee (Office of Parking Services). University of Nevada, Las Vegas, 2013-present.

Meeting Organization

- Co-organizer of symposium for 252nd American Chemical Society National Meeting. Philadelphia, PA, August 2016.
- Co-organizer of symposium for 2016 Biennial Conference on Chemical Education. Univeristy of Northern Colorado, CO, July 2016.

Outreach/ Community Service

- Guest Speaker (SISTEM: Student Interactions with STEM), University of Nevada, Las Vegas, 2017.
- Graduate Student Mentor (Center for Academic Enrichment and Outreach), University of Nevada, Las Vegas, 2017.
- Presidential Student Ambassador (Office of the President), University of Nevada, Las Vegas, 2015-2016.
- Volunteer (Shade Tree Women's Shelter), Las Vegas, Nevada, 2015-present.