Exploring problem solving strategies on multiple-choice science items: Comparing native Spanish-speaking English language learners and mainstream monolinguals

Rachel Rae Kachchaf

University of Colorado at Boulder, rkachchaf@gmail.com

Follow this and additional works at: http://scholar.colorado.edu/educ_gradetds

Part of the Educational Assessment, Evaluation, and Research Commons, First and Second Language Acquisition Commons, and the Science and Mathematics Education Commons

Recommended Citation

EXPLORING PROBLEM SOLVING STRATEGIES ON MULTIPLE-CHOICE
SCIENCE ITEMS: COMPARING NATIVE SPANISH-SPEAKING ENGLISH
LANGUAGE LEARNERS AND MAINSTREAM MONOLINGUALS

by

RACHEL RAE KACHCHAF

B. A., College of Charleston, 2004
M. A., New York University, 2006

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado at Boulder in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Education
2011
This thesis entitled:
Exploring problem solving strategies on multiple-choice science items: Comparing native Spanish-speaking English language learners and mainstream monolinguals
written by Rachel Rae Kachchaf
has been approved for the Department of Education, University of Colorado, Boulder

Kathy Escamilla, Ph.D.

Janette Klingner, Ph.D.

Bhuvana Narasimhan, Ph.D.

Guillermo Solano-Flores, Ph.D. (Chair)

Lucinda Soltero-González, Ph.D.

The final copy of this thesis as been examined by the signatories and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

HRC protocol # 0108.25
Kachchaf, Rachel Rae (Ph.D., Education)

Exploring problem solving strategies on multiple-choice science items: Comparing native Spanish-speaking English language learners and mainstream monolinguals

Thesis directed by Professor Guillermo Solano-Flores

The purpose of this study was to compare how English language learners (ELLs) and monolingual English speakers solved multiple-choice items administered with and without a new form of testing accommodation—vignette illustration (VI). By incorporating theories from second language acquisition, bilingualism, and sociolinguistics, the study was able to gain more accurate and comprehensive input into the ways students interacted with items. This mixed methods study used verbal protocols to elicit the thinking processes of thirty-six native Spanish-speaking English language learners (ELLs), and 36 native-English speaking non-ELLs when solving multiple-choice science items.

Results from both qualitative and quantitative analyses show that ELLs used a wider variety of actions oriented to making sense of the items than non-ELLs. In contrast, non-ELLs used more problem solving strategies than ELLs. There were no statistically significant differences in student performance based on the interaction of presence of illustration and linguistic status or the main effect of presence of illustration. However, there were significant differences based on the main effect of linguistic status. An interaction between the characteristics of the students, the items, and the illustrations indicates considerable heterogeneity in the ways in which students from both linguistic groups think about and respond to science test items.

The results of this study speak to the need for more research involving ELLs in the process of test development to create test items that do not require ELLs to carry out significantly more actions to make sense of the item than monolingual students.
Acknowledgements

I am indebted to the children and teachers in this study for allowing me into their classrooms. This dissertation would not have been possible without you.

I am especially grateful to Dr. Guillermo Solano-Flores for his guidance and support throughout this journey. I am fortunate to have had the opportunity to learn from you. I am also very thankful to the members of the committee, Dr. Kathy Escamilla, Dr. Janette Klingner, Dr. Bhuvana Narasimhan, and Dr. Lucinda Soltero-González for your mentoring on this dissertation and beyond. Additionally, the support of the entire EECD community has been astonishing, I am grateful to all of you.

Finally, I would be nowhere without the support of my family who have been with me every step of the way. My father and his faith in education have been especially motivating. When planning for a year, plant corn. When planning for a decade, plant trees. When planning for lifetime, educate people.

Funding for this dissertation was provided by the National Science Foundation, Award No. DRL 0822362. The opinions expressed here do not necessarily reflect those of the funding agency.
Table of Contents

List of Tables ................................................................................................................................. viii
List of Figures ................................................................................................................................... x

Chapter 1: Introduction ..................................................................................................................... 1
  Statement of the Problem ........................................................................................................ 3
  Purpose Statement ................................................................................................................... 5
  Significance ........................................................................................................................... 6
  Conceptual Framework ........................................................................................................... 7
  Research Questions ............................................................................................................... 20

Chapter 2: Review of the Literature ............................................................................................... 22
  Problem Solving Strategies ................................................................................................... 23
  Cognitive Validity in Educational Testing ............................................................................ 29
  Testing Accommodations for ELLs ..................................................................................... 44
  Summary .............................................................................................................................. 51

Chapter 3: Methodology .................................................................................................................. 53
  Strategy of Inquiry ................................................................................................................. 53
  Setting and Participants ........................................................................................................ 53
  Researcher Role ..................................................................................................................... 56
  Item Selection ........................................................................................................................ 57
  Illustration Development ...................................................................................................... 59
  Verbal Protocol Development ............................................................................................... 60
  Interview Procedures ............................................................................................................. 60
  Coding System ...................................................................................................................... 63
List of Tables
Table 1: *Demographics of Emerging Bilinguals and Mainstream Students* ........................................... 54
Table 2: *Items Used in the Investigation* .................................................................................................. 58
Table 3: *Coding System: Making Sense* .................................................................................................. 65
Table 4: *Criteria for Classification of Science Terms* ........................................................................... 69
Table 5: *Coding System: Problem Solving Strategies* ........................................................................... 71
Table 6: *Coding System: Vignette Illustration Use* ............................................................................... 75
Table 7: Item 1: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* ........ 84
Table 8: Science Terms not Understood by ELLs ......................................................................................... 86
Table 9: Item 2: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* ........ 98
Table 10: Item 3: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 104
Table 11: Item 4: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 111
Table 12: Item 5: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 116
Table 13: Item 6: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 121
Table 14: Item 7: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 129
Table 15: Item 8: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 134
Table 16: Item 9: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* .......... 140
Table 17: Item 10: *Frequencies of Most Common Codes by Linguistic Status and Coding Category* ........ 146
Table 18: *Demographics: Students using Spanish* .................................................................................. 151
Table 19: *Use of Spanish by Interview Section* ..................................................................................... 151
Table 20: Spanish by Interview Section on Common Items ....................................................................... 152
Table 21: *Use of Spanish and Problem Solving Strategies* .................................................................... 154
Table 22: *Use of Spanish and Problem Solving Strategies: Common Items* ......................................... 154
Table 23: ANOVA Results for the Frequency Differences of Codes for Making Sense by Linguistic Status and Presence of Illustration .................................................................................................................. 162
Table 24: Making Sense: Frequencies of Most Common Codes by Linguistic Status, Regardless of Presence of Illustration .................................................................................................................. 165
Table 25: Making Sense: Frequencies of Most Common Codes by Presence of Illustration and Linguistic Status .................................................................................................................................. 168
Table 26: Making Sense: Frequencies of Most Common Codes by Presence of Illustration and Linguistic Status: Items 1 and 2 .................................................................................................................................. 171
Table 27: Mean Number of Actions by Linguistic Status .......................................................................................................................... 174
Table 28: Mean Number of Actions by Illustration and Linguistic Status .......................................................................................................................... 176
Table 29: Mean Number of Actions to Make Sense of Items by Grade and Linguistic Status ........ 177
Table 30: Mean Number of Actions to Make Sense of Items by ELL Classification ................. 177
Table 31: ANOVA Results for Differences in the Frequencies of Codes for Problem Solving Strategies .................................................................................................................................. 180
Table 32: Frequencies of Most Common Codes for Problem Solving Strategies: Items 3-10, Regardless of the Presence of Illustration .................................................................................................................. 182
Table 33: Frequencies of Most Common Codes for Problem Solving Strategies by Linguistic Status and Presence of Illustration .................................................................................................................. 185
Table 34: Items 1 and 2: Frequencies of Most Common Codes for Problem Solving Strategies ... 188
Table 35: Frequencies of Most Common Codes for Use of Vignette Illustration by Linguistic Group and Item .................................................................................................................................. 192
Table 36: Student Performance Across Illustrated and Non-illustrated Versions of Items: Rounded Percentages .................................................................................................................................. 195
Table 37: ANOVA Results for Student Performance by Linguistic Status and Presence of Illustration .................................................................................................................................. 195
List of Figures

Figure 1. Conceptual Framework .......................................................................................................... 8
Figure 2. Linguagram ............................................................................................................................ 17
Figure 3. Item 1: Seedling .................................................................................................................. 84
Figure 4. Combination tree: Bilingual students ............................................................................... 94
Figure 5. Combination tree: Monolingual students ......................................................................... 95
Figure 6. Item 2: Moon ...................................................................................................................... 97
Figure 7. Item 3: Alexander Fleming ............................................................................................... 103
Figure 8. Item 4: Stopwatch ............................................................................................................. 110
Figure 9. Item 5: Helium Balloon .................................................................................................. 111
Figure 10. Item 6: Pinewood ............................................................................................................. 120
Figure 11. Item 7: Magnesium Oxide .............................................................................................. 128
Figure 12. Item 8: Platypus .............................................................................................................. 133
Figure 13. Item 9: Cold Region ........................................................................................................ 138
Figure 14. Item 10: Earth’s Temperature ......................................................................................... 145
Figure 15. Making sense: Correctly answered Item 3-10 ................................................................. 163
Figure 16. Making sense: Incorrectly answered items 3-10 ............................................................. 163
Figure 17. Making sense: ELLs and Non-ELLs .............................................................................. 166
Figure 18. Making sense: Illustrated vs. non-illustrated items ........................................................ 167
Figure 19. ELLs: Making sense by presence of illustration ............................................................. 169
Figure 20. Non-ELLs: Making sense by presence of illustration ..................................................... 169
Figure 21. Illustrated items: ELLs vs. non-ELLs ............................................................................ 169
Figure 22. Non-illustrated items: ELLs vs. non-ELLs ..................................................................... 169
Figure 23. ELLs vs. non-ELLs: Making Sense of Item 1 ................................................................. 172
Figure 24. ELLs: Making sense of Item 2 ........................................................................................ 173
Figure 25. Problem solving strategies: Correctly answered Items 3-10 ........................................ 181
Figure 26. Problem solving strategies: Incorrectly answered Items 3-10 ....................................... 181
Figure 27. Problem solving strategies: ELLs vs. non-ELLs ............................................................. 183
Figure 28. Problem solving strategies: Illustrated vs. non-illustrated .......................................... 184
Figure 29. Problem solving strategies: ELLs ............................................................................... 186
Figure 30. Problem solving strategies: Non-ELLs ....................................................................... 186
Figure 31. Problem solving strategies: Illustrated items ................................................................. 186
Figure 32. Problem solving strategies: Non-illustrated items ....................................................... 186
Figure 33. Problem solving strategies: Item 1 ............................................................................. 189
Figure 34. Problem solving strategies: Item 2 ............................................................................. 190
Figure 35. Vignette illustration use: Items 3-10 ......................................................................... 191
Figure 36. Vignette illustration use: Item 1 .................................................................................. 193
Chapter 1

Introduction

In today’s classrooms, English Language Learners (ELLs)\(^1\) are the fastest growing student population (Flynn & Hill, 2005). Currently, over 11% of students in the U.S. are classified as ELLs (Pearson, 2006) and some researchers predict that by the year 2030 over 40% of all students will come from homes where English is not the main language spoken (Thomas & Collier, 2002). Furthermore, 24 states, including Colorado, have seen over 100% growth in their ELL population since 1994, including some states, such as Kentucky and South Carolina, which have experienced over 400% growth (Payan & Nettles, 2006).

This rapid increase in linguistic diversity calls attention to the need for more research involving ELLs, also referred to here as emerging bilinguals. While the population of students classified as ELLs speaks more than 450 languages, 79% speak Spanish as a first language (Dalton, Sable, & Hoffman, 2006). Therefore, focusing on Spanish speaking ELLs can help improve the education of the majority of linguistically diverse students.

Adding to the complexity of the educational challenges that ELLs face, current legislation requires that all students, regardless of English proficiency, participate in state-mandated standardized testing after living in the United States for one year. Students may take the test in their native language only if they were born outside the U.S. and have lived in the country for less than three years. This instance applies to less than half of ELLs, as the vast majority of students classified as ELLs are born in the country and are typically required to participate in standardized testing in English (Kohler & Lazarín, 2007).

---

\(^{1}\) I prefer to use the term *emerging bilinguals* to refer to the bilingual participants in this study. However, due to terminology used by policies, schools, and previous research, I may also use the term English language learner (ELL). Please note that these terms will be used interchangeably throughout the discussion.
Requiring emerging bilinguals to take tests in English completely disregards the linguistic characteristics of bilingual students and contradicts research showing that it may take them five to seven years to reach a proficiency level in English at which they can benefit from teaching and be fairly tested in English (e.g., Hakuta, Butler, & Witt, 2000). Additional research estimates that it can take up to ten years for some students to become proficient in academic English (Collier, 1995). Some experts (e.g., Valdés & Figueroa, 1994) argue that current testing practices lack any theoretical foundation and question whether ELLs should even be expected to participate.

In spite of these challenges in the testing of ELLs, researchers (e.g., Abedi, Hofstetter, & Lord, 2004) point out that excluding ELLs from standardized testing could have extremely negative consequences. The inclusion of ELLs in state mandated tests often determines distribution of funding for these students and it monitors the adequacy of their overall education. Exclusion of ELLs from standardized testing could result in a loss of funding and other important resources. Given the critical role that testing plays in ELL education, much improvement of assessment practices is needed if states and schools are to make fair, important decisions about students, teachers, and schools based on standardized test scores.

Motivated by the urgent need to improve testing conditions for linguistically diverse students, the Illustrations in the Testing of English Language Learners (ITELL) project explores the possibility of using vignette illustrations as a form of testing accommodations for ELLs (Soalno-Flores, 2008a). Testing accommodations for ELLs include any modification that changes the way in which a test is administered with the intent to reduce the cognitive demands posed on ELLs resulting from testing them in their second language (Butler & Stevens, 1997; Durán, 2008; Rivera, Collum, Wilner, & Sia, 2006). Later described in detail, vignette illustrations are defined as images added to multiple-choice items as a form of visual support that
preserves the original text of the item (Solano-Flores, 2010a; 2010b). This four-year National Science Foundation funded project was informed by the fields of semiotics, cognitive psychology, linguistics, and socio-cultural theory to more appropriately approach the development of this new form of testing accommodation. The ITELL investigation attended to four aspects of using accommodations that are not typically acknowledged by current practice. These aspects included: (1) to provide appropriate conceptual foundations to support the development of vignette illustrations as accommodations, (2) to minimize the effect of limited English proficiency on ELLs’ performance without changing the constructs measured by the original items, (3) to create an accommodation that is easy to implement, and (4) to develop an accommodation that is cost-effective.

Due to the fact that, with a few exceptions, illustrated items are not addressed by research in testing, one goal of the ITELL project was to inform current practice about how this new form of testing accommodation can reduce the role of English proficiency in emerging bilinguals’ performance. Using vignette illustrations as a testing accommodation is promising because it does not require any reworking of the original text of the items (therefore reducing the possibility of changing the construct measured) and the fidelity of implementation is not dependent on administrators’ skills.

As a sub-study of the ITELL project, this dissertation focuses on comparing aspects of answering multiple-choice items by native Spanish speaking ELLs and monolingual mainstream students taking items both with and without this new testing accommodation.

**Statement of the Problem**

Current assessments are inaccurate measures of what linguistically diverse students know (Solano-Flores, 2006; 2008a). Many scholars attribute this inaccuracy to the exclusion of
linguistically diverse populations during the test development and piloting phases of large-scale, 
standardized assessment (Abedi, Hofstetter, & Lord, 2004; Valdés and Figueroa, 1994). Because 
test development typically excludes emerging bilinguals, item writers have no empirical 
evidence about any differences that exist between bilinguals and monolinguals responding to 
multiple-choice items. Rather, piloting is typically carried out with mainstream monolingual 
students, and it pays little attention to the linguistic and cultural factors that play a major role in 
how ELLs’ solve test items. Exclusion of emerging bilinguals is problematic because, as Kopriva 
(2000) points out, linguistically diverse students differ in fundamental ways from their 
mainstream counterparts. Not considering, or not investigating these differences when 
developing test items can result in tests that do not accurately measure what ELLs know.

Recent attention to the need for improved testing conditions for linguistically diverse 
students has focused on the use of various types of accommodations. Testing accommodations 
for ELLs are intended to reduce language as a source of construct irrelevant variance, or 
minimize the effect of limited English proficiency on students’ performance. While research has 
yielded, at best, mixed results about their effectiveness (Abedi & Hefri, 2004; Kieffer, Lesaux & 
Rivera, 2009; Sireci, Li & Scarpati, 2003), accommodations are the primary tool that states 
utilize to justify the inclusion of ELLs in large-scale tests (Rivera, Collum, Wilner, & Sia, 2006).

Furthermore, studies investigating the use of a specific accommodation seldom give 
empirical data in support for their use; rather, the support is often theoretical, based on expert 
judgment, or student performance (e.g., test scores). No study has investigated the specific 
thinking processes of ELLs responding to test items with and without a given accommodation. 
More research is needed to investigate the detailed interaction between emerging bilinguals and
test items to inform the testing community about the ways in which these students interact in assessment situations.

The scant research that investigates the testing of linguistically diverse students often uses what Grosjean (1989) calls fractional views of bilingualism. These fractional views fail to accurately consider the full linguistic repertoire of bilinguals and to adequately incorporate their knowledge of two languages into all aspects of their research designs. Although previous studies (e.g., Bailey, 2007; MacSwan, Rolstad, & Glass, 2002) have revealed inaccuracies about the inferences drawn from tests using fractional views of language, research has yet to investigate these issues in-depth when students take multiple-choice tests with and without a specific accommodation.

**Purpose Statement**

The purpose of this mixed methods study is to utilize cognitive interviews to explore how native English-speaking monolinguals and native Spanish-speaking bilingual students classified as ELLs solve multiple-choice science items with and without vignette illustrations. The analysis focuses on two aspects: (1) the ways that emerging bilinguals and mainstream monolinguals make sense of multiple-choice test items, and (2) the problem solving strategies that emerging bilinguals and mainstream students use. I identify commonalities and differences in the inferred thinking processes that mainstream monolingual students and bilingual Latino students use to respond to accommodated and non-accommodated versions of items. Because no study has previously investigated these issues, I seek to provide empirical evidence about ELLs’ thinking processes when solving items. The main focus is on how ELLs make sense of items and the strategies they invoke to solve items. With this focus in mind, I also examine the specific ways in
which ELL and non-ELL students use vignette illustrations when responding to multiple-choice science items.

**Significance**

This study accomplishes three goals: 1) to inform the education community about the ways that bilingual students make sense of multiple-choice items and the reasoning they use to solve the items, 2) to inform the educational assessment community about important differences in the thinking processes between monolinguals and bilinguals requiring that bilinguals be included in the process of test development, and 3) to provide an example of how the educational assessment community can appropriately incorporate bilinguals into the beginning phases of test development.

First, to my knowledge, no study exists that investigates the specific thinking processes of bilingual students classified as ELL taking multiple-choice science items. The information from this study provides invaluable insight for the education community. By systematically investigating the strengths and challenges emerging bilinguals face in testing situations, teachers can have a better understanding of the experiences their bilingual students go through when they solve multiple-choice items. Not only is this information useful for bilingual or ELL teachers, but it is useful for content teachers who have ELLs in their classrooms. Furthermore, teachers can use this knowledge to inform how they use formative assessments in the classroom.

Second, while the assessment community has noted the importance of including bilinguals in the process of test development, this study provides detailed information about why their inclusion is essential. A systematic comparison of the experiences solving items of monolinguals and bilinguals gives item developers an idea of what differences exist between these two groups of students. This comparison informs all aspects of test development, especially item generation,
by illustrating difficulties ELLs experience interpreting items and determining if the strategies items writers intended are in fact those that ELLs use to solve items.

Finally, this study illustrates how the assessment community can incorporate theories of bilingualism and sociolinguistics into all the phases of developing tests and testing accommodations to more appropriately address the needs of linguistically diverse students. By considering both of emerging bilinguals’ languages during the entire study, test developers can have a better understanding of how these issues can be addressed during all stages of test development.

**Conceptual Framework**

August and Hakuta (1997) state that research in second language acquisition depends on the researcher’s definition of language. As previously noted, fractional views of bilingualism fail to consider important differences between monolinguals and bilinguals and have tainted the knowledge base of research involving bilinguals (Grosjean, 1985; 1989). Solano-Flores (2009) argues that current approaches in the assessment of ELLs are based on deterministic views of language that force decisions about these students to be based on student demographic variables that provide limited valuable insight about ELLs. Not only is it important to view ELLs from a holistic view of bilingualism that considers both of students’ languages at all times; it is also essential to recognize the extreme heterogeneity that exists among bilinguals. This requires researchers to acknowledge that each bilingual has his/her own unique system of language use. To this end, Solano-Flores (2009) posits that a view of ELL testing as a stochastic process is more appropriate because it recognizes that many factors describing and addressing students’ language proficiency are uncertain or beyond control.
Here, I present a three-part conceptual framework, shown in Figure 1, that provides a comprehensive lens to investigate the thinking processes of Spanish-speaking emerging bilinguals taking multiple-choice science items with and without illustrations. At left, the figure shows an interaction between the holistic view of the bilingual individual and the sociolinguistics factors that influence a bilingual individual’s language use. The first layer, a holistic view of bilingualism, recognizes cognitive differences between monolinguals and bilinguals. The second layer of the framework, sociolinguistic aspects that influence bilinguals’ language use, describes the multiple factors that shape the ways bilinguals choose to use language across situations. By first considering these two aspects, I am able to more comprehensively and accurately investigate how emerging bilinguals solve multiple-choice science items.

**Figure 1.** Conceptual framework

**Characteristics of the bilingual individual.** I utilize a broad definition of the bilingual individual: any person who routinely uses two languages (Gumperz, 1973). I find it necessary to
categorize participants based on one criterion described by Valdés and Figueroa (1994) as fundamental for any researcher conducting research with bilingual individuals—being an elective or a circumstantial bilingual.

Valdés and Figueroa (1994) describe elective bilinguals as those individuals who choose to learn a second language (L2). These individuals may seek out foreign language classes and travel to or temporarily live in a foreign country where the language is spoken. The individual’s first language (L1) holds the highest prestige in society. This type of bilingualism is additive—the person is adding a language to his/her repertoire. Because there are individual motives for learning the second language, the L2 is considered to be a characteristic of the individual, not the group. For example, a person from the U.S. who wants to learn German may seek out foreign language classes and travel to Germany to learn the language. His/her native language (English) holds the highest prestige in the home country.

Valdés and Figueroa (1994) distinguish elective bilinguals from circumstantial bilinguals. They describe circumstantial bilinguals as those individuals who are required to learn a second language to survive or participate in society due to their circumstances. For example, immigration, colonization, shifting borders, and other political reasons may require individuals to learn an L2 to succeed in mainstream society. Their L1 is not the language of the majority. Although the individual can choose not to learn the L2, this decision typically holds many negative consequences for their future as the L1 has little prestige in greater society. This type of bilingualism is considered to be subtractive bilingualism because it often results in the loss of the L1 in favor of gaining the L2. Because there are societal reasons for learning it, the L2 is considered to be a characteristic of groups of people, rather than an individual. Examples of circumstantial bilinguals include American Indians or Mexican immigrants living in the U.S.
The participants in this study are circumstantial bilinguals, and, as Bialystok (2002) notes, all their cognitive processes are intertwined. Therefore, bilinguals’ cognitive processes fundamentally differ from monolinguals because they constantly draw from both languages, even when activities are carried out in one language. These processes include memory, perception, and language. Emerging bilinguals have the knowledge of both these languages entangled with all other cognitive processes. Bialystok further describes language proficiency as the bilingual’s ability to extract linguistic regularities through multiple interactions with the world around them. The bilingual then accrues these regularities to build knowledge. In this model of cognition, interaction is key to knowledge construction. It also requires continual consideration of students’ L1 and L2 at all times to accurately interpret a bilingual’s thinking processes. As Grosjean (1985; 1989) notes, a bilingual is not equal to two monolinguals in one, rather they draw from different linguistic resources and face different challenges in communicating.

The point that a bilingual’s knowledge of two languages is intricately intertwined with all other cognitive processes leads to Cummins (1981) theory of common underlying proficiency. He explains that bilinguals have one underlying system from which they draw to communicate in either language. Therefore, bilinguals must be able to use both of their languages to have full access to this underlying system. Grosjean (1989) describes bilinguals as living in a variety of contexts that range in a spectrum from completely monolingual situations, where only one language is used, to completely bilingual situations, where two languages are continually mixed. Grosjean emphasizes that, at any given time, a bilingual can be at any point along this spectrum using only one language or a combination of both languages. In addition, two bilinguals might be at different points on the spectrum when communicating in the same context. Furthermore, some bilinguals may prefer monolingual situations where they use only one language whereas other
bilinguals may prefer bilingual contexts where they continually incorporate both languages into their communications.

Pushing this notion of bilingualism further, bidirectionality explains the bidirectional relationship between a bilingual’s two languages. That is, not only does the L1 support and influence the use of the L2, but knowledge of the L2 also supports and influences the use of the L1. Dworin (2003) emphasizes that it is important to consider this bidirectional relationship because it amplifies the ability for bilinguals to draw from two cultural and linguistic resources. Bidirectionality also acknowledges that language is both a sociolinguistic and psycholinguistic in nature as it involves both cognitive and social factors. Accessing both languages has been shown in previous investigations to be crucial for bilinguals to successfully carry out various academic tasks (e.g., Escamilla, 2000; Moschkovich, 2007).

While these theories highlight similarities in bilinguals’ cognitive processes, there is also great heterogeneity. This heterogeneity stems from a plethora of reasons, including the fact that there is no uniform starting point for each bilingual (Gass & Selinker, 2001). Furthermore, all bilinguals differ in numerous aspects at the time they began learning a second language, including, but not limited to, demographic characteristics (e.g., age when acquisition of English began and birthplace). Bilinguals also differ in their abilities in each of their languages, described by Mackey (1968) as the degree to which they use both languages.

First, all individuals vary in the degree to which they use both languages across the four modes of language: reading, writing, listening, and speaking. Second, these abilities change across different domains or registers. For example, one bilingual can be more comfortable speaking about sports in his L1 and more comfortable writing about science in the L2. The same bilingual can be comfortable writing about many domains in his L1 but very comfortable
listening and speaking about them. This point is further discussed in theories relevant to the sociolinguistic aspects influencing bilinguals’ language use, but here I recognize that bilinguals differ in the degrees to which they use language and that abilities in each language are not static conditions, but rather constantly evolving states (Gumperz, 1973).

Other aspects that play a role in describing bilinguals’ language repertoire can be broken down into three categories: psychological, cultural, and social (Chin & Wigglesworth, 2007). Psychological aspects included the order in which students learned both their languages. Some students may have learned their languages sequentially or simultaneously (Romaine, 1995). Simultaneous bilinguals describe children born in the U.S. who started acquiring two languages between the ages of 0-5 (Baker, 2006). The majority, over 50%, of emerging bilinguals in Grades 6-12 are simultaneous bilinguals (Capp, Fix, Murray, Ost, Passel, & Herwantoro, 2005). Many of these young bilinguals are still in the process of developing their L1 at the time they begin learning the L2. Other bilinguals are classified as sequential bilinguals because they are considered fluent in their L1 when they begin to learn their L2 (Katz, Low, Stack, & Tsang, 2004).

Additional psychological aspects describing bilinguals include anxiety, personality, and motivation (Gass & Selinker, 2001; Romaine, 1996). Anxiety is the pressure each individual feels when constructing or maintaining a positive appearance to others. Personality characteristics are the extent to which bilinguals are introverted or extroverted and influence the amount of risks they are willing to take with respect to language. Next, each bilingual differs in his/her overall intelligence level and motivation to learn languages. Finally, bilinguals vary in their familiarity with the cultural beliefs and norms of dominant society and the degree to which
those beliefs and norms differ from those of his/her native culture (Abella, Urrutia, & Schneyderman, 2005).

Cultural aspects contributing to the heterogeneity among bilinguals include how the native culture views bilingualism and the L1 or L2 individually. This is not to say that all individuals from the same culture share the same view on bilingualism. For instance, a Mexican-American family may view bilingualism positively while another may hold negative views about it. There are also different expectations about language use and traditional ways in which bilinguals are socialized to use language across cultures and within cultures (Echevarria & Short, 2002; Heath, 1983; Katz et al., 2004). In addition, parents use different strategies to help their children develop languages and use languages differently at home (Baker, 2006; Romaine, 1995; 1996). Some parents may encourage the use of both L1 and L2 at home, while others encourage the use of only the L1 or L2.

The degree to which knowledge from a bilingual’s L1 transfers to L2 also differs across languages (Gass & Selinker, 2001). Many studies have shown that knowledge of language from the L1 can directly transfer to learning the L2 (see August & Hakuta, 1997). However, at times, knowledge of the L1 can cause interference when learning the second language. Mackey (1968) defines interference as any instance in which a bilingual incorrectly applies norms of language use from one language to the other.

Social factors describing bilinguals’ background include their varying levels of socioeconomic status, although the majority of ELLs are from low socioeconomic groups (Payan & Nettles, 2006). In addition, the attitude toward both their first language and their abilities in the second language held by their schools, government, and society can differ (Romaine, 1995). Some students attend English-only schools, an environment that may reflect an overall negative
view of bilingualism and use of the L1. Other students attend bilingual programs that may reflect a more positive view of bilingualism. However, bilingual programs differ by type (e.g., dual immersion and transitional).

All of these factors contribute to heterogeneity among bilinguals. In addition, each of the factors discussed here affects each bilingual differently (Valdés & Figueroa, 1994). For example, even when students have similar school experiences, levels of motivation, and personality types, these factors can impact each individual’s language use in a different way. To this point, the educational research community has yet to accurately acknowledge this heterogeneity and incorporate it into testing practices (Solano-Flores, 2009).

**Sociolinguistic view of bilinguals’ language use.** The second component of the framework incorporates sociolinguistic principles to describe how bilinguals use language across contexts. First, I again acknowledge the extreme heterogeneity that exists among bilinguals. As described in the previous section, bilinguals’ linguistic repertoires are the result of their interactions with the world (Bialystok, 2002). Coulmas (2005) adds that each speaker continually makes choices about how to use language in a given situation. Reyes (2004), among many others, explains that the choices bilinguals make are influenced by numerous contextual factors, including the topic or the interlocutor’s characteristics. Furthermore, an individual’s linguistic repertoire changes over time (Grosjean 1998; Gumperz, 1973). All of these sociolinguistic aspects result in the fact that each bilingual has his/her own unique linguistic repertoire—a repertoire that differs from those of all other bilinguals. As Dworin (2003) and Chin and Wigglesworth (2007) note, there are multiple paths to bilingualism.

This heterogeneity among bilinguals is often ignored in research. Valdés (1992) states that typically, bilinguals are poorly grouped together and inadequately described. Kopriva (2007)
urges the assessment community to consider background variables such as age of acquisition of English, length of time in the U.S., and literacy in native language. Kopriva explains that by collecting these variables researchers may be able to determine trends across ELLs with similar backgrounds.

At the same time, Dworin (2003) cautions researchers not to overgeneralize based on generic background characteristics. Valdés and Figueroa (1994) also warn that broad demographic categories do not allow the researcher to make any predictions or assumptions about how language will be used. That is, one cannot assume that any two bilinguals will use language in the same way based on the simple facts that they have spent the same amount of time in the U.S. and both began learning English at the same age. Dworin warns that even descriptors such as native language may be inappropriate for young bilinguals simultaneously learning two languages. While it is important to recognize that background characteristics are needed to adequately describe an individual, this information must not be used to make assumptions about an individual’s language use to describe his/her thought processes (Prosser & Solano-Flores, 2010).

Not only do the backgrounds of ELLs greatly vary, so does their use of languages across contexts. Here, it is helpful to refer back to the spectrum provided by Grosjean (1989) that describes bilinguals as living in a span from completely monolingual to completely bilingual events. Valdés and Figueroa (1994) present a similar spectrum describing a bilingual’s preference for language, given a specific situation. At one end of the spectrum is the speaker’s preference to use only one language, represented as A. At the other end of the spectrum is the speaker’s preference to use only the second language, B. As one approaches the middle of the spectrum, the preferences for communication begin to incorporate two languages. The center
represents situations where there is very little preference for using one language over the other. Below is a simplified example of this spectrum, as illustrated by Valdés and Figueroa:

\[
\begin{array}{cccc}
A & Ab & aB & B \\
\end{array}
\]

For example, a young bilingual communicating with his grandmother at a wedding about his favorite food may prefer to use only language A, represented on the spectrum by point A. The same young bilingual that is discussing his favorite video game with a cousin may show only a very slight preference for language A over language B, represented as Ab. Finally, the same bilingual who is discussing a homework assignment with his bilingual friend at an English-only school might have a strong preference for language B over language A, represented as aB. It is important to acknowledge that each bilingual may make different choices of how to use language in the same situation.

Building upon the spectrum described by Valdés and Figueroa (1994), I also recognize that bilinguals’ use of both languages differently not only differs across contexts, but also across language modes. Solano-Flores and Li (2008) provide a linguagram to represent the relationship between a bilingual’s two languages across each mode: reading, writing, listening, and speaking. Figure 2 shows their example of a linguagram:
The authors use this linguagram to conceptually represent the fact that bilinguals’ proficiency levels are different across each mode in each language. For example, Student 1 is much more proficient in the four modes of the first language than the four modes in the second. In contrast, Student 3 appears to be very proficient speaking and listening in both languages but is less proficient in reading and writing in either language. Not only do bilinguals differ in the ways that they use each language according to their context, but they also differ in their proficiencies across modes for a given context. This means that the linguagram shown in Figure 2 will change according to the specific situation in which the bilingual is communicating.

In addition to preference of using one language over another, some bilinguals prefer to mix languages or code-switch. For the purposes of this study, I utilize Gumperz’s (1973) definition of code-switching: the alternate use of two or more languages in an utterance or conversation.
code-switching, such as the purpose for switching languages (e.g., McClure, 1981; Poplack, 1980; Reyes, 2004; Zentella, 1981), this is outside the scope of this study. The current focus is recognizing that bilinguals may choose to mix languages.

Reyes (2004) found that modeling language use encouraged participants to use whichever language they preferred in academic settings. By switching languages throughout the introduction the researchers gave sociolinguistic contextual clues to the participants that using either language was an acceptable way to communicate. Zentella (1981) points out that use of both languages is a very delicate situation. She discovered that minor dialectal differences among interlocutors can discourage a bilingual from code-switching. However, she also found that young bilinguals typically speak as they are spoken to. Both researchers describe that young bilinguals have different preferences for language use.

Fishman (1965) describes many sociolinguistic factors that influencing bilinguals’ language use during interactions, including group membership and role-relations. Group membership refers to characteristics determined by a combination of factors such as age, sex, ethnicity, and native language, among many others. Speakers who share group membership most likely communicate in different ways than speakers who do not share group membership. Role-relations describe the power structure between speakers and also influence the ways individuals interact using language.

**Problem solving strategies.** Theories of problem solving differ in the number of stages and steps they identify. However, most researchers agree on two main aspects of problem solving: item comprehension and problem space (Leighton & Gokiert, 2005; Pretz, Naples & Sternberg, 2003). The first category, item comprehension, refers to the ability of the student to make sense of the text of the item. As Polya (1973), points out, this initial stage is crucial for successful
problem solving. Here, it is important to acknowledge that ELLs and monolinguals may have very different experiences when making sense of an item in English due to cognitive differences discussed in earlier sections of the conceptual framework.

After making sense of the item, a student enters what Leighton and Gokiert (2005) call the problem space. Leighton and Gokiert explain that this metaphorical “space” refers to students invoking specific strategies and knowledge to arrive at a solution. While previous research has investigated the difference in the problem spaces between expert and novice solvers (e.g., Chi, Glaser, & Farr, 1988), research has yet to focus on the differences in the problem spaces of emerging bilinguals and mainstream students.

The first two parts of this framework, utilizing a holistic view of bilingualism and acknowledging sociolinguistic factors that influence individuals’ language choices, allow for the more accurate and comprehensive investigation of how students carry out these two steps for problem solving. The intent of this framework is to minimize the measurement error of exploring students’ thinking processes. In psychometrics, measurement error is the result of construct irrelevant variance. Construct irrelevant variance occurs when student performance varies due to constructs others than those intended to be measured (Messick, 1989). As discussed later in detail, previous research (e.g., Abedi, 2006) argues that language is often a source of construct irrelevant variance in the testing of emerging bilinguals. However, research has yet to identify the precise role that language plays in bilinguals’ performance on standardized tests.

Although the current study gives weight to qualitative methods, psychometrics plays a central role for two reasons. First, understanding the specific thinking processes of emerging bilinguals is crucial to create tests that more accurately measure what linguistically diverse students know. Without conceptual consideration of the aspects described in the first two parts of
the framework, comprehensive information about bilinguals thinking processes would not be obtained. Second, cognitive interviews are commonly used during test development to inform the creators about how a test can be improved. Therefore, my study directly relates to the development of assessments for bilinguals.

Solano-Flores (2006; 2008a) urges the assessment community to view testing situations from a sociolinguistic perspective by considering the interaction between student and test item as communication. The same approach applies to the use of cognitive interviews to inform assessment development. Sudman, Bradburn, and Schwartz (1996) utilized a sociolinguistic perspective in their study with monolinguals. They investigated the interaction between the student, the test item, and the interviewer. This approach provided them with a comprehensive view of the event. In this study, the interaction between student, item, and interviewer is also considered, keeping in mind the importance of recognizing two languages in the cognitive processes of bilinguals, as shown by Escamilla (2000) and Moschkovich (2007).

**Research Questions**

The main research question for this study is:

How do native Spanish-speaking ELLs and native English-speaking mainstream students compare in their inferred thinking processes when they solve multiple-choice science items with and without vignette illustrations?

To answer this main question, I investigate three sub-questions:

1. What are the similarities and the differences in the ways that native Spanish-speaking emerging bilinguals and native English-speaking mainstream students make sense of multiple-choice science items with and without vignette illustrations?
2. What are the similarities and the differences in the strategies that native Spanish-speaking emerging bilinguals and native English-speaking students use to solve multiple-choice science items with and without vignette illustrations?

3. What are the similarities and the differences in the ways that native Spanish-speaking emerging bilinguals and native English-speaking mainstream students use vignette illustrations when solving multiple-choice science items?

4. How do bilingual students draw from Spanish to solve multiple-choice science items with and without vignette illustrations?
Chapter 2

Review of the Literature

Overview

Due to the fact that scant research exists in the qualitative exploration of emerging bilinguals’ experiences interacting with test items, this literature reviews draws from various fields to inform the current study. The review is divided into three major sections: problem solving strategies, cognitive validity in educational testing, and issues in the testing of emerging bilinguals.

In the first section, I briefly review relevant literature on the problem solving strategies of monolingual mainstream students. Next, I review the few studies that have investigated problem solving among bilinguals. Although not all of these studies come from the field of educational testing, they inform the current study by focusing on differences between bilinguals and monolinguals.

The second section discusses issues of validity in educational assessment. After providing an overview of approaches to validity, I discuss how a specific tool, the verbal protocol, has been used to investigate issues of validity in educational testing with both monolingual and bilingual students.

Finally, the third section discusses issues in the testing of emerging bilinguals. I focus on major issues in obtaining accurate measures of what linguistically diverse students know. Within this category, I discuss previous attempts in the field to gain more accurate measures, including the use of testing accommodations. This section concludes by focusing on a specific accommodation, the vignette illustration, the focus of the ITELL project.
Problem Solving

Problem solving and monolinguals. There is a vast research base investigating various aspects of students’ problem solving strategies on different types of academic tasks. Many studies have focused on differences between experts and novices in a given content area (e.g., Chi, Glaser, & Farr, 1988). While a complete review of problem solving is outside the scope of this paper, attention is given to studies that investigate the different phases of problem solving as well as specific strategies students use to solve problems. As previously noted, researchers vary in the number of steps they identify in the process of solving a problem, but they commonly identify two essential steps: understanding and problem space.

Many researchers argue that the first step in solving a problem, understanding, is the most important of all steps (e.g., Nathan, Kintsch, & Young, 1992; Polya, 1973; Pretz, Naples, & Sternberg, 2003). As Polya (1973) points out, “the worst may happen if the student embarks upon computation or constructions without having understood the problem.” (p. 5). However, understanding the task at hand is not only critical for the participant, but also for the item writers. Exploring the details of how students understand test items is a crucial step into investigating whether or not an item functions as item writers expected (Leighton & Gokiert, 2008; Messick, 1989).

Cummins, Kintsch, Ruesser, and Weimer (1988) investigated the role of text comprehension and successful problem solving with elementary students. They gave 18 mathematical word problems to 38 Grade 1 students who were required to either solve the problem and recall the question or recall the question and then solve the problem. Using a linguistic developmental view, they considered certain word problems to be more difficult to solve because they “employ linguistic forms that do not readily map onto children’s existing conceptual knowledge” (p. 407).
The authors found that there was indeed a strong relationship between text comprehension factors and successful problem solving. They identified various types of miscomprehensions which lead the researchers to conclude that the misinterpretations of items were systematic. They conclude that poor linguistic knowledge results in poor problem solving while robust linguistic knowledge leads to successful solution attempts.

**Problem solving and bilinguals.** Morales, Shute, and Pelligrino (1985) investigated the problem solving of 61 Grade 3 and 50 Grade 5 and 6 Mexican-American students from low socioeconomic backgrounds on subtraction and addition word problems. Half the students were from bilingual classrooms, and half were from monolingual classrooms. The authors found that older students (Grades 5 and 6) were more successful and systematic problem solvers than younger students (Grade 3). The differences in students’ abilities to solve problems within each grade level were associated with the nature of the task rather than the language of the problem (Spanish or English). This finding was consistent for students who attended bilingual and monolingual classes. However, the authors failed to provide detailed information about the ways students made sense of items or the strategies they used to solve. Instead, they associated differences in student performance with the demands a task placed on students.

Duran (1985) tested 209 bilingual Puerto Rican college students on their logical reasoning abilities. He investigated the relationship between reading comprehension in both Spanish and English as a predictor of logical reasoning in both languages. After collecting data on reading comprehension and logical reasoning in both languages, he found that reading comprehension and logical reasoning ability were strongly associated. He also found that reading speed was a strong predictor of logical reasoning ability. While Duran found that students typically performed better on the English version of the tests, individuals who scored equally on reading
comprehension in both languages performed equally on the logical reasoning tests in both languages. Although this study provides some insight into the link between reading abilities and reasoning abilities, Duran failed to consider whether the tests used to obtain scores of students reading comprehension were accurate measures.

Secada (1991) investigated the semantic structures of word problems to analyze the influence of specific grammatical features on bilinguals’ ability to correctly answer problems. He categorized items according to their semantic structure and tested 45 Hispanic students in Grade 1 from low and middle socioeconomic status in the Chicago area. Using the Language Assessment Scales (LAS) to test for the degree of bilingualism, Secada found that performance was consistent across languages, but participants tended to be slightly better solvers in English. Secada concluded that the Hispanic students were more similar than different from mainstream students when solving problems because both groups of students performed similarly on items with varying semantic structures. That is, the items that were the most difficult for mainstream students were also the most difficult for Hispanic students. Secada argued that the LAS did not accurately predict problem solving abilities of Hispanic students because it does not incorporate the varying semantic structures that are directly associated with bilinguals’ problem solving abilities. However, the results of this study must be cautiously interpreted as Secada does not clearly define his population. Hispanic students attended both monolingual and bilingual programs, an aspect that was not accounted for in the analysis. Furthermore, it appears that Secada uses a deficit view of bilingualism, focusing on the skills that students lack rather than focusing on those skills students possess.

Mestre (1988) conducted two investigations to analyze how language interacted with the problem solving performance among bilinguals in Grade 9 and undergraduate college students.
In the first study, six Grade 9 Spanish and English bilinguals and eight monolinguals solved algebra word problems. In the second study, college undergraduate engineering students read a statement using negations and then needed to select affirmative statements that they believed conveyed the first statement. Using verbal protocols in both of these studies, Mestre concluded that there are four challenging aspects of language for bilinguals in problem solving: general language proficiency, technical language proficiency, mathematical use of language, and symbolic language. He points out that the role of language is often down played when discussing the steps to problem solving. He argues that this lack of recognition is unfortunate because successful problem solvers must be linguistically precise. Often students who misinterpret the item continue on to correctly solve the item based on the initial misinterpretation. Therefore, students have mathematical ability but are unable to show it due to misunderstanding aspects of the item. For example, he found that misinterpreting prepositions resulted in students incorrectly translated word problems to the mathematical representation of $6H = P$, which should have been represented as $H = 6P$. This study was one of the first to systematically examine whether items accurately measure what linguistically diverse students know. Unfortunately, the very small sample sizes limit the generalizations of these findings.

Chamot, Dale, O’Malley, and Spanos (1992) investigated the implementation of a curriculum for upper elementary students that focused on teaching ELLs the academic language of mathematics and science. Per the curriculum, teachers were trained to emphasize a specific sequence of the steps to solve problems: understand the question, find the data, make a plan, solve the problem, and check the solution. The authors argued that, for ELLs in upper elementary grades, following this sequence is crucial to correctly solving problems. The authors found that high implementation classrooms had a higher percentage of high ability students completing
steps in the proper sequence, and correctly solving the problem. The authors state that these results imply that it is more important for ELLs than for monolinguals to follow a rigid sequence of steps when solving problems. However, the authors assumed that high implementation classes resulted in using the specified sequence of steps and they did not investigate whether students actually utilized these steps when solving problems.

Jimenez, Garcia, and Pearson (1996) investigated the reading strategies of 14 Grade 6 and 7 students. As one of the few studies that utilized a holistic view of bilingualism, the authors identified eight successful English/Spanish bilingual readers, two marginally successful English/Spanish bilingual readers, and three successful monolingual English readers. The authors found that all eight successful bilingual readers knew to search for cognates, while four of the successful bilingual readers routinely translated from one language to other to make meaning of the text. When comparing the bilingual students reading in Spanish and English, they found that students had less prior knowledge in Spanish, but monitored their comprehension more when reading in Spanish. The authors argued that these are strategies teachers can teach their bilingual students to become better readers. While this study utilized a holistic framework and had interesting results, their small sample size prohibits one from generalizing from these results.

Bialystok and Majumder (1998) investigated the advantages bilinguals have over monolinguals when solving nonlinguistic problems. While previous studies have not given much attention to the proficiency level of participants, these authors tested their participants for their language abilities in L1 and L2. They studied 71 middle class children from three different cultural backgrounds in Grade 3. Participants were placed in three different groups according to language proficiency: monolingual, unbalanced bilinguals and balanced bilinguals. By giving participants four different tasks, the authors concluded that balanced bilinguals perform better on
tasks that require a higher control of attention. Their results lead them to argue that all studies
must control for language proficiency to accurately analyze and interpret results. However, I
agree with the authors who caution that the three groups of students included in their study were
from very different cultural backgrounds, and this could have affected results.

Bernardo (2002) investigated the relationship between native language and students’ abilities
to understand problems in both languages. He conducted interviews with 92 bilinguals, half
classified as native Filipino-speaking, and half classified as native English-speaking. All
participants were in Grade 2 and took 18 arithmetic word problems. After examining the role of
comprehension in problem solving, Bernardo found that students better understood (and
performed better on) problems written in their native language. Unfortunately, results from this
study fail to provide much insight because Bernardo fails to give specific details about the
specific linguistic features of items and the role of these features in students’ experiences solving
problems.

In their qualitative study, Parvanenhnezad and Clarkson (2008) investigated 16 Iranian
fourth- and fifth-grade students’ language use in Australia. Students completed 10 open-ended
mathematics word and symbolic problems. Students used a checklist to keep a record of which
language they used to solve each problem. Parvanehnezad and Clarkson found that those
students who had high English language proficiency typically had a high mathematical
competency. In addition, only two students reported not using both Persian and English to solve
the problems. These results add to the research that indicates bilinguals must be allowed to use
their full linguistic repertoire to complete academic tasks (e.g., Escamilla, 2000; Moschkovich,
2007). However, their sample sizes were small and, therefore, there are serious limits to the
extent to which the results can be generalized.
The common thread among the studies included in this section was the relationship between language ability and student performance. However, most of these investigations did not utilize a framework that considers the holistic and bidirectional relationship between bilinguals’ two languages. Therefore, the results must be interpreted with this point in mind. While these studies have specifically focused on differences between monolinguals and bilinguals, none of them have explored the detailed experiences of emerging bilinguals in a testing situation responding multiple-choice science items.

**Cognitive Validity in Educational Testing**

Validity in educational assessment is defined as the degree to which evidence and theory support interpretations of test scores entailed by proposed uses of tests (AERA/APA/NCMA, 1999). That is, can the conclusions drawn from a given test be sufficiently supported by empirical and theoretical evidence? A test in and of itself is not valid or invalid, but rather the uses of or interpretations drawn from the test results are valid or invalid. Validity is considered to be the most fundamental consideration when developing tests (AERA/APA/NCME, 1999). Perhaps this is because the act of measurement uses limited samples of observations to draw important conclusions about student ability (Kane, 2006).

In his discussion of validity, Messick (1989) describes three types of validity: content validity, criterion-related validity, and construct validity. Content validity refers to whether or not a given test assesses the intended subject matter (e.g., general science knowledge or more specifically, the cycles of life). Criterion-related validity compares the student’s test score to an external variable (criteria) that represents a direct measure of the construct being measured (e.g., comparing a test measuring general science knowledge to a student’s science grade in science class). Criterion-related validity can be predictive or concurrent. Predictive validity refers to the
ability of a current measure of a student’s ability to predict performance on a future measure of the same construct. Here, it is important to note that, although many standardized tests are strong predictors of mainstream students’ abilities, they often underestimate the potential of culturally and linguistically diverse students (Klingner, Blanchett, & Harry, 2007). Concurrent validity refers to the degree to which a measure represents the student’s ability, as reflected by another, previously validated measure, at the time he/she completes the test. Finally, construct validity refers to the ability of a test to measure the intended target, rather than aspects outside this intended target, such as language proficiency if the target is science knowledge. The field of educational measurement has termed any variance attributable to factors other than the construct intended to be measured as construct-irrelevant variance.

Early discussions of validity centered on the types of evidence that can be used to create a validity argument in educational testing. These early discussions considered criterion validity to be the gold standard (Kane, 2006). However, Messick (1990) called for an integrated approach noting that there was danger in only providing one type of evidence to support validity. He noted that almost all content validity arguments were expert judgments of whether the test content covered the domain. Evidence supporting these claims rarely addressed any interaction between the test and the student. For example, Messick points out that it is possible, based on expert judgment, to produce a test that is deemed to assess science content when in fact it also measures reading comprehension (p. 11). He also notes that a test scores may vary due to differences in its format (e.g., multiple-choice or constructed response) rather than the content tested. Thus, presenting evidence to support content validity is not enough on its own. Likewise, merely presenting evidence to support only criterion-related validity or only construct validity is deemed insufficient.
Rather than outlining specific types of evidence that can be used to create a case for validity, more recently discussions of validity, such as those given by Kane (2006), describe validity as building an argument which can be divided into two parts. These parts are: the interpretive argument (a general framework for the interpretation and use of test scores) and the validity argument (an evaluation of the interpretive argument). Creating an acceptable case for validity includes more than addressing content, criterion, and construct validity. It also requires incorporation of several types of evidence to create what Kane calls a “mini-theory” to support the argument for validity. This argument for validity is evaluated on clarity, internal consistency and plausibility of inferences and assumptions. The resulting argument should serve three functions: (1) provide a framework for test development by indicating what assumptions should be met, (2) provide a framework for the validity argument, and (3) provide a basis for evaluating the validity argument as a whole.

**Validity arguments in the case of testing linguistic minorities.** In the case of testing linguistic minorities, within the past decade, aspects of language proficiency have become an integral part of validity arguments. Due to the fact that any test that uses language is partially a measure of the test taker’s language skills (AERA/APA/NCME, 1999), special attention must be given to language in the case of testing emerging bilinguals. The accuracy of a test score is compromised by the fact that it is influenced, to a large extent, by a student’s limited proficiency in the language of testing. Therefore, test norms based on native speakers of English should not be used with individuals whose first language is not English (AERA/APA/NCME, 1999, p. 91). The standards specify that current testing practice should be designed to reduce threats to validity and that these validity arguments must entail evidence for linguistic subgroups just as the evidence used for mainstream students.
Garcia and Pearson (1994) describe three types of bias from which traditional testing suffers. The first, norming bias, results from the small samples of linguistic minorities that are included in the populations of students used to norm tests. The authors point out that traditional approaches to creating tests require examination of the statistical properties of test items. Any item that does not behave as predicted is considered problematic and revised or removed from the final version of the test. Linguistic minorities are typically low-scoring, and Garcia and Pearson explain that this step in test development necessarily means that many items on which linguistically diverse students do well on disappear from the final version of the test. That is, if many low-scoring students correctly answer a test question, this is an unpredicted occurrence based on the statistical properties of the item. That test question is subsequently removed without further investigations into the reasons why low-scoring students were able to accurately answer the question. Garcia and Pearson argue that this type of process in test development fails to account for the complexity of student and item interaction.

The second type of bias Garcia and Pearson (1994) describe is content bias, which refers to the fact that current testing procedures reflect the dominant cultures’ expectations for language use and shared knowledge. Although early test developers attempted to design tests that were culture free, this was found to be impossible because tests are artifacts of culture (Cole, 1999) and they are therefore necessarily tied to the culture. Indeed, the authors show that these “culture free” tests did not decrease the correlation between socioeconomic status and IQ performance (p. 345). Garcia and Pearson explain that the tests reflected cultural values of the mainstream such as language skills (e.g., discourse). Garcia and Pearson also reviewed state reforms that intended to remove or decrease such bias, but the authors deemed them unsuccessful.
The third type of bias includes linguistic and cultural aspects. Linguistic bias refers to English vocabulary or ways of speaking that are more familiar to mainstream students than to linguistic minorities. The cultural bias includes aspects of test taking that are characteristic of mainstream students but not linguistically diverse students. The authors give the example of speededness, the rate at which different students take tests, as an example of cultural bias. Students from different cultures may spend more or less time answering items. Accurately estimating the amount of time students need to complete the test requires including diverse students in all phases of test development. The authors link these types of bias to consequential validity and state that, while it is impossible to avoid the social nature of tests, these differences become extremely problematic when tests are used to make decisions about placement into classes of linguistically diverse students.

Abedi (2006) has shown that current assessments suffer from weak validity arguments. That is, the inferences drawn from student performance on tests are not accurate conclusions about what linguistically diverse students know. In his many studies on the testing of ELLs, he has found the linguistic complexity of test items to be a source of construct-irrelevant variance. He outlines many aspects of the language of testing that pose more difficulty for ELLs than for native English speakers. These aspects include the use of complex sentence structures, conditional clauses, long noun phrases, prepositional phrases, negation, and passive voice. He argues that these factors partially explain why ELLs score lower on standardized tests than non-ELLs.

Solano-Flores (2006) points out the complex nature between two parts of language of testing: dialect and register. Dialect refers to a variation of a language that is characteristic of the users of that language, whereas register refers to a variation of a language that is determined by its use (p.
He explains that ELLs’ performances on tests are very sensitive to dialect variations because these students are still developing their first and second languages. In addition, students must master the testing register, typically used only in assessment situations, if they are to perform well on standardized tests. When students’ knowledge of a dialect and the register used on a specific test match up, there is linguistic alignment; however when these aspects do not match up, there is misalignment. Solano-Flores uses a theoretical perspective that describes how students can handle misalignment within limits: if there are too many instances of mild misalignment or a few but severe instances of misalignment, the student’s performance will most likely be affected.

Unfortunately, to this point, research has relied heavily on the theoretical reasoning to explain aspects of ELLs’ test performance. That is, in validity arguments, most researchers use theory to support their claims but have yet to conduct in-depth investigations into the specific details explaining ELLs’ test performance. Kopriva (2001) suggests incorporating the use of cognitive labs to collect empirical evidence to explore cognitive aspects of students taking test items. She states that these labs should focus on ELLs’ understanding of what a particular item asks them to know and the specific role construct irrelevant factors play in their interpretation. In addition, cognitive labs should elicit information from students to explain why they arrived at a particular response. In order to gather accurate insight, Kopriva advises that participants represent the full range of English proficiencies. Then, analysis should compare patterns of understanding for students of different proficiencies.

**Verbal protocols and monolinguals.** Cognitive interviews have been used since the early 1900’s in various fields including cognitive psychology (e.g., Newell & Simon, 1972), criminal
investigation (e.g., Memon & Bull, 1991), and survey development (e.g., Sudman, Bradburn, & Schwartz, 1996).

This section focuses on the use of cognitive interviews in educational assessment. Cognitive interview is a term that can refer to any interview used with the intent to uncover the mental processes of the participant (Zucker, Sassman, & Case, 2004). For the purposes of this study, I am concerned with cognitive interviews that utilize verbal protocols, a form of instrument that Ericsson and Simon (1993) describe as providing insight into the cognitive processes a participant uses for a specific task.

Verbal protocols consist of concurrent and/or retrospective reporting. Ericsson and Simon (1993) define concurrent reporting as states of heeded information that are directly verbalized. This type of vocalization occurs as the participant completes the task at hand, in this case a multiple-choice science item. Ericsson and Simon define retrospective reporting as trace accounts of information taken from short-term memory or retrieved from long-term memory. This verbalization can occur anytime after the task is complete, typically immediately following completion of the task.

Ericsson and Simon (1993) recommend that researchers use a combination of concurrent and retrospective reporting to achieve the most comprehensive insight into the participant’s cognitive processes. Paulsen and Levine (1999) echo this suggestion and explain that concurrent reporting provides insight into the mental processes as they occur in real time. Retrospective reporting provides participants’ accounts of what processes he/she used to complete the task, therefore shedding light on his/her understanding of the experience solving the test question. Although these two types of reporting often result in similar information, they can offer slightly different insight into the mental processes. Leighton (2004) also notes that concurrent reporting identifies
the knowledge and skills actually used to solve items, and retrospective reporting can clarify or elaborate any ambiguities about this reasoning stated in the concurrent reporting. These three types of reporting allow for triangulating the information students provide on their reasoning with the intent to gain the most comprehensive information about students’ experiences solving multiple-choice science items.

Some researchers have challenged Ericsson and Simon’s claims that verbalizations do not alter the naturally occurring thinking processes of the participant. Smagorinsky (1998) analyzed the use of verbal protocols through the lens of Culturally Historical Activity Theory. He argues that the simple presence of another person (the interviewer) necessarily influences the participant, thus changing the naturally occurring thinking processes. However, Ericsson and Simon (1998) stand behind their claims, presenting evidence from over 30 sources to support their argument. They explain that if researchers follow the directions and do not interject prompts other than “keep talking,” the participants’ thinking processes will not be influenced.

Norris (1990) was among the first researchers to investigate the usefulness of verbal protocols in educational assessment. He analyzed whether participation in verbal protocols influenced student performance on multiple-choice items. He divided 343 Canadian high school students in Grades 10-12 into five groups: no verbal protocol, concurrent reporting only, retrospective reporting only, and two different conditions of follow-up questioning. Students in each group took the same multiple-choice test assessing one’s ability to judge the credibility of reports of observations. For each item, the students read two eye-witness accounts and judged which of the reports was more credible. He found no significant differences in the scores of each group of students. The results of his study showed that student performance was not affected by participation in any of the different interview situations. Based on these findings, he concluded
that all types of cognitive interviews could be used with students taking multiple-choice assessments without impacting their ability to perform. In addition, he found that students in each of the five groups provided very similar information. He concluded that verbal protocols were a valid tool to inform test development about the thinking processes of students.

Although Norris (1990) found that similar information was elicited from students in the various types of reporting, a later study by Taylor and Dionne (2000) compared the types of knowledge accessed through concurrent reporting and retrospective reporting. In their study, 36 participants carried out concurrent and retrospective reporting. The authors found that concurrent reporting yielded more information on the specific actions that participants carried out while retrospective reporting yielded more conditional knowledge. Conditional knowledge referred to the participant’s ability to state when and where a specific strategy was utilized, and how it relates to other strategies in terms of difficulty. The results supported the author’s argument that using a combination of these two types of reporting allowed access to a broader spectrum of insight into participants’ knowledge of how they solved problems. It also revealed a more accurate picture of the various strategies used by participants with different levels of expertise. In addition to eliciting different types of information, these complimentary types of reporting can also be used to triangulate data by comparing information participants provide in each type of reporting.

Other studies utilized verbal protocols to investigate the reasoning students use on assessments that do not utilize a multiple-choice format. Baxter, Elder, and Glaser (1996) used cognitive interviews to investigate Grade 5 students’ reasoning on performance tasks. They found that students’ performances could be placed in three groups: consistently high, consistently low, and intermediate. The researchers also found that the information elicited from
students through the use of verbal protocols provided extremely useful insight into specific aspects of student reasoning on the performance tasks. This information could then be used to inform instruction about students’ knowledge of the content and how they can reach deeper levels of understanding. For example, by pinpointing details of students’ misconceptions, teachers could more effectively teach the topic to overcome any misconception.

Hamilton, Nussbaum, and Snow (1997) used cognitive interviews to clarify results of large-scale testing by collecting qualitative information to explain the quantitative data provided by students’ test results. By interviewing 41 high school student volunteers, they investigated the reasoning students used on three different types of items: multiple-choice, constructed-response, and performance task. As the authors point out, some educators criticized multiple-choice items for requiring simple recall of factual knowledge; however, the researchers found that multiple-choice items required three dimensions of ability. They called these dimensions: quantitative science, spatial-mechanical reasoning, and basic knowledge. The quantitative science dimension included chemistry items, items requiring mathematical computations, or both. The spatial-mechanical reasoning dimension included items that required visualization, extracting information from diagrams, or both. The basic knowledge and reasoning dimension included items that required students to recall factual knowledge from various domains of science. While it was commonly believed that constructed response and performance tasks provided students the opportunity to use higher level thinking skills, the researchers found that students were often able to arrive at the correct answer to items in these formats without using higher level thinking skills. The authors used these results to caution test developers against making assumptions about the thinking processes a particular format an item requires. In a subsequent study, Ayala, Shavelson, Yin, and Schultz (2002) found that mathematical performance assessments can also tap higher
level skills. They found that the performance tests they investigated involved the same three dimensions: quantitative science, spatial-mechanical reasoning, and basic knowledge and reasoning.

Ruiz-Primo, Shavelson, Li, and Shultz (2001) used concurrent and retrospective reporting to investigate the different techniques 152 high school students used to create concept maps in science assessments. At the micro-level, they coded verbalizations, focusing on specific sentences or utterances, to describe how students defined, compared, and justified their actions when creating a concept map. At the macro-level, they focused on the entire content of student verbalizations to identify ways in which students planned and used strategies. By capturing activities at both levels they were able to describe various cognitive activities involved in creating concept maps. Ruiz-Primo and her colleagues found that their coding framework allowed them to distinguish between different types of reasoning that students used to create concept maps.

Not only have educational researchers utilized verbal protocols to investigate aspects of cognitive processes of students across various types of academic tasks; they have also used them to investigate students’ interpretations of items. Norris, Leighton, and Phillips (2004) state that verbal protocols are a valuable tool to investigate whether students focus on the target construct. They conclude that eliciting students’ thinking processes can be used to investigate whether they allocate adequate attention to the critical aspects of the item.

Leighton and Gokiert (2005; 2008) did just this in their study of 54 Grade 8 and Grade 11 students. These researchers found that using unfamiliar or polysemic words often resulted in students misunderstanding the item. In their interviews, words with multiple meanings, ambiguous wording, and incomplete contextual clues were often a source of confusion or
misinterpretation. I agree with the authors who argue that similar investigations should be carried out during test development to inform how item writers create items.

**Verbal protocols and linguistically diverse students.** With few exceptions (e.g., Prosser & Solano-Flores, 2010), research using cognitive interviews has not focused on how the verbal protocol should be modified for appropriate use with a new population of students. Rather, researchers either simply apply how the tool is used with monolinguals to bilinguals with little consideration for differences in the communication styles of the different linguistic groups.

Although their study did not utilize verbal protocols, Solano-Flores and Li (2009) investigated differences among three different cultural groups in interviews about the ways in which the students related test items to their personal experiences. They found that students differed in two areas: length of student response and level of detail of their descriptions. However, they also found that rigorous coding procedures ensure the validity of the information obtained for each cultural group.

Winter, Kopriva, Chen, and Emick, (2006) utilized verbal protocols to investigate the performance of 156 Grade 3 and Grade 5 students to explore their access to item content. The authors define access as, “the interaction between construct-irrelevant item features and personal characteristics that either permits or inhibits student response to the targeted measurement content of the item” (p. 268). That is, there might be some aspects that the items measured that were not initially intended to be measured. These aspects may help or hinder students to correctly answer the item. The authors argue that, while access to test content is an issue for all students, it is much more complex for ELLs than for mainstream students due to language and cultural factors. They explored student thinking processes in three stages of problem solving: comprehension of task, formulation of the solution, and articulation of the solution. Their results
showed that students who comprehended the item correctly were more likely to select an appropriate solution strategy. Furthermore, students who were provided with item modifications that improved their ability to comprehend the item were much more likely to select the appropriate solution strategy and correctly answer the item.

Kopriva, Cameron, Carr, Wright, and Bauman (n.d.) used verbal protocols to investigate how 58 students in Grades 4-9 interacted with a dynamic computer-based science items. Dynamic computer-based science items are administered via a computer and utilize animations to set up the context of the item. In addition, students could use the computer mouse to roll over text in the item to highlight support that attempts to clarify the text’s meaning. As part of the process of development, the researchers used verbal protocols to investigate how students interacted with items administered via the computer program. They focused on whether the animation, support, or item features increased or decreased students’ abilities to access the content of the item. While this work is still in progress, preliminary results show that there is a need to balance item text and other visual elements on the screen. This balance may be attained by standardizing the formatting of items administered via the computer program. When this goal is achieved, the authors argue that use of such a computer program could be an effective way to assess ELLs’ knowledge.

Maritiniello (2008; 2009) used verbal protocols in attempts to disentangle the relationship between language skills and mathematics proficiency in the mathematics assessment of ELLs. She examined differential item functioning (DIF), which shows if students from different populations (e.g., gender) have different probabilities of correctly answering the same question after differences in ability is controlled for. She examined the differences in performance between ELLs and their mainstream counterparts, focusing on whether the magnitudes of DIF
were associated with the nonmathematical linguistic complexity of the items. Next, Martiniello interviewed 24 Grade 4 ELLs using verbal protocols to capture how they used non-linguistic representations to scaffold their understanding of the items. She found that syntax, lexicon, cultural references, and layout were potential sources of difficulty for bilinguals when comprehending test items. However, other non-linguistic components, such as diagrams, were shown to attenuate these differences.

Ercikan, Arim, Law, Domene, Gagnon, and Lacroix (2010) also used verbal protocols to further explain statistical results obtained using DIF. The authors point out that, while expert review is the most commonly used method to identify aspects of items that can lead to DIF, these arguments are not empirically supported. The researchers conducted think aloud interviews with students to investigate if the items’ surface features identified by expert reviewers as sources of DIF were supported by students’ verbalizations of their thinking processes. Thirty-six English speaking, seventh-graders and twelve French speaking, Grade 7 and 8 students took standardized test items from a Canadian national assessment that was concurrently developed in English and French. Next, the authors used the verbalizations from think aloud interviews to assist the interpretation of the DIF results.

Through this analysis, they identified four aspects that may have resulted in items functioning differently for students. First, they looked at student understanding (and misunderstanding) of the item text, and found differences in difficulty of certain words in one language versus the other. Second, they examined if students perceived the item to be difficult. The authors found that this information was not helpful in examining DIF. Third, students were able to point out what aspects of the question were helpful, but this information was only useful if students were able to identify specific words, rather than just vague statements, such as “the
item was clear” or “the table helped” (p. 33). Finally, the authors found the most informative insight to be when students identified confusing or difficult aspects of the item.

By comparing information from think-aloud protocols to the expert review of items, Ercikan et al. found an overlap in 10 of the 20 items investigated. That is, of the 20 items identified to function differently for the bilinguals and monolinguals, the expert panel accurately identified why the items functioned differently for only 10 of the items. The authors conclude that expert review panels alone are not a sufficient way of identifying problematic aspects of items. This information must be combined with the information gleaned from students’ cognitive interviews to achieve comprehensive understanding of how items function differently across groups of students.

These previous investigations utilizing cognitive interviews with bilinguals were only somewhat helpful in the ways researchers should utilize verbal protocols with bilinguals. Specifically, they failed to provide detailed information about the precise communicative practices of bilinguals during the cognitive interview. Due to this gap in the literature, I conducted a study to analyze the language use of bilinguals in cognitive interviews (Prosser & Solano-Flores, 2010). Results showed that the majority of ELLs participated in the interview in English only, but that a significant amount (about 30%) chose to communicate in a combination of English and Spanish, regardless of their background characteristics. That is, students who used Spanish were classified at various levels of English proficiency and had lived in the country for varying amounts of time. Therefore, using verbal protocols with ELLs requires permitting all students to use either language.
Testing Accommodations for ELLs

As previously mentioned, a major concern in the testing of ELLs is minimizing measurement error due to the test taker’s proficiency in the language of testing. One approach to minimizing this error is the use of testing accommodations. For an accommodation to be considered valid, it should increase the scores of ELLs but not affect scores of native English-speakers (Sireci, Scarpati, & Li, 2005).

Researchers have found a total of 75 different testing accommodations for ELLs allowed by states across the country (Kieffer, Lesaux, Rivera, Francis, 2009; Rivera, Collum, Wilner, & Sia, 2006). Although the goal of testing accommodations is to provide linguistic support, many of the allowed accommodations are taken from the field of special education and provide no linguistic support (Rivera & Collum, 2006). In fact, over half of the states in the U.S. organize ELL accommodations within the taxonomy traditionally used to classify accommodations for students with disabilities failing to recognize fundamental differences between these two groups of students.

Furthermore, there is very little empirical support to justify using accommodations. As Butler and Stevens (1997) note, the need to include ELLs is immediate but there is a great lack of empirical evidence to support the use of any accommodation. Current support is given theoretically or based on student test scores. There is also a lack of support for the way testing accommodations are selected for ELLs. In their national study of how accommodations were assigned, Rivera, Collum, Wilner, and Sia (2006) found that only 27 of the 47 states use language-related variables to make decisions about which accommodations ELLs received. While some states provide instructions explaining how accommodations should be used (e.g., Colorado Department of Education, 2008), others provide no guidance at all. Finally, another
issue with accommodations is that there is no way to track the fidelity of implementation of accommodations across states. For example, the accommodation, oral script requires that the administrator of the accommodation read the text aloud to the student. The administration of this accommodation can vary depending on the individual characteristics of the person providing the accommodation to the student(s). Current practice does not systematically investigate the consistency of how this accommodation is administered.

Rather than reviewing all studies investigating testing accommodations for ELLs, I discuss two reviews of the literature that show comprehensive results about previous research on accommodations. In addition to these reviews, I include a review of studies examining the latest accommodation being used with ELLs, computer administrated tests.

In their review of the literature on the effectiveness of testing accommodations, Abedi, Hofstetter, and Lord (2004) included 11 studies investigating the effectiveness of various accommodations. They found that the most common testing accommodation used with ELLs was extra time given to students to complete the test. Other accommodations included the use of a native language test, use of a published dictionary, use of a glossary (or customized dictionary), oral administration of test items, and modified English. The authors caution that research should not seek to find one accommodation that will work for all ELLs. Instead, they offer four aspects for consideration when selecting accommodations: effectiveness, validity, differential impact, and feasibility. That is, one must consider if the accommodation: (1) improves test scores for ELLs, (2) does not affect the scores of mainstream students, (3) does not affect students’ scores from different backgrounds differently, and (4) is relatively easy to implement. The researchers found that only two accommodations met these criteria: modified English and use of a glossary (customized dictionary). However, these accommodations were only effective in a few of the
reviewed studies. These mixed results led the authors to advise large-scale testing to include ELLs in the process of test development from the beginning, rather than as an afterthought. In addition, they recommend that future research examine the effectiveness and validity of accommodations for ELLs based on background variables such as length of time in the U.S. and language spoken at home.

Keiffer, Lesaux, Rivera, and Francis (2009) conducted a meta-analysis on the effectiveness and validity of testing accommodations for ELLs. Their meta-analysis of 11 total studies found that there was an overall lack of evidence for the effectiveness of any testing accommodation currently used. The authors analyzed the effectiveness of five accommodations: English glossary, bilingual glossary, simplified English, native language test, and dual language test. Consistent with the findings of Abedi, Hofstetter and Lord (2004), they found that English glossaries had an overall positive effect on ELLs’ performances. Although this effect was statistically significant, the effect size was small. Therefore, there is little empirical support that any of the accommodations currently in use improve ELLs’ scores. The authors also investigated whether non-ELLs’ receiving accommodations showed improvement in their test scores. Kieffer et al. concluded that there was “little cause for concern” (p. 1183). That is, the scores of mainstream students did not significantly increase when taking an accommodated version of the test.

Since the publication of these reviews, researchers have investigated the effectiveness of a new accommodation: computer testing. Abedi (2009) investigated 666 Grade 4 and 643 Grade 7 ELLs’ and mainstream students’ performances on items that were administered via computers. For this accommodation, students could gloss over various words to expose a pop-up box with a definition of the word. Based on previous research that indicates linguistic complexity has a
negative impact on ELLs’ performance, the researchers rated items on a scale of 0-4 for
linguistic complexity. Grade 4 students participated in one of five situations: computer
accommodation, extra time, customized dictionary, small group administration, or no
accommodation. Results showed that students who took computer administered tests scored
statistically significantly higher than students in the non-accommodated group. Students
receiving extra time also scored statistically significantly higher than the non-accommodated
group. There were no statistically significant differences between mainstream students taking
any of the accommodated versions of the test versus those students receiving no accommodation.
Grade 4 students scored significantly higher regardless of the linguistic complexity rating of the
items. Grade 8 students were divided into three groups: computer administered test, customized
dictionary and no accommodation. Results showed that only ELLs accommodated by the
computer administered test had statistically significantly higher scores than ELLs receiving the
customized dictionary and ELLs with no accommodation. Similar to the results with Grade 4
students, mainstream students showed no significant score differences in any of the three groups.
However, when Abedi broke down items based on their linguistic complexity, he found that the
computer accommodation only resulted in statistically significant score differences between
ELLs receiving the accommodation and not receiving the accommodation on items classified as
linguistically complex. In addition to test score differences, Abedi found that ELLs glossed over
twice as many words to access their definitions. Abedi concludes that computer based testing
shows potential for a successful testing accommodation for ELLs. He also acknowledges that the
exact nature of how this new accommodation works is unknown. It is possible that taking tests
via computer is seen as a privilege or is more fun for students, which contributes to the
difference in scores.
Overall, studies investigating the use of current testing accommodations have not provided clear evidence that those accommodations reduce construct irrelevant variance in the testing of emerging bilinguals. In addition, no study has investigated bilinguals’ thinking processes when using a particular accommodation. Conclusions about the effectiveness of accommodations were based on test scores and theoretical arguments. Perhaps investigating the specific experiences of ELLs using a testing accommodation could inform the field about how their use can be improved.

**Vignette illustrations.** A new form of testing accommodation currently in development is the vignette illustration. Vignette illustrations refer to images that are created to add to test items originally developed without illustrations (Solano-Flores, 2010a; 2010b). Rather than test items that directly refer a student to an image, vignette illustrations are not necessary to solving the problem. Because this type of accommodation does not alter the original text of an item, it has potential to minimize measurement error due to language without changing the construct measured by the item. A vignette illustration is defined by three characteristics: (1) the illustration provides a simple, concrete representation of one of the components of the text of the item, (2) the text of the item does not refer the test taker to the illustration, (3) the test of the item provides all the information needed to understand it and respond to it, even if the illustration is removed.

The methodology for creating vignette illustrations is informed by the fields of semiotics, cognitive science, sociolinguistics, and socio-cultural theory. Semiotics informs the examination of the relationship between text and image. Cognitive science informs the inferred mental activity elicited by this relationship. Integrating knowledge from sociolinguistics and sociocultural theory allows careful examination of how linguistic and cultural influences play a
role in shaping this mental activity (Solano-Flores, 2008a). Vignette illustrations are developed through collaboration of a multidisciplinary team including science teachers, bilingual teachers, scientists, and illustrators. The involvement of these various experts allows the team to develop illustrations that can attend to the linguistic challenges of ELLs while accurately representing science concepts.

A review of related literature revealed one dissertation that investigated the use of images as a form of testing accommodations, and found that no previous study has investigated the use of vignette illustrations. Shanahan (2006) compared student performance in three testing situations: items with images, items with graphic organizers, and items with no accommodation. 86 urban Grade 5 ELLs took a 20-item multiple-choice life science test. She found no statistically significant differences between the non-accommodated and accommodated test items. However, her study included no discussion of how images were created other than that they focused on “contextual information.” Likewise, the graphical organizer accommodation was only briefly described as showing hierarchical relationships between science concepts related to the concepts assessed by the item. This lack of conceptual foundation makes interpretations of her results very difficult. Furthermore, her study was purely quantitative and did not involve any analysis of the interaction between the item and student thinking processes.

Literature from the field of cognitive psychology provides valuable information about the role of images in the ways students learn content. Mayer, Heiser, and Lonn (2001) investigated the effect of including static images as compared to animated narrations. They investigated if the use of multimedia could help teach scientific explanations. The authors based their investigation on previous literature that supports the information delivery hypothesis: the idea that learning through two modalities is always better than one because each modality is a delivery system for
information. In their study with undergraduate college students, the researchers conducted four experiments to investigate differences in retention and transfer of knowledge. These experiments engaged students to learn about how lightning works in a variety of conditions: (1) animated video with an on-screen text summary vs. animated video with an on-screen text summary with extraneous details, (2) no text in the animation vs. summary text in animation vs. full text in animation, (3) animated video with no text vs. interesting but conceptually irrelevant video interdispersed throughout video, and (4) animated video with no text vs. interesting but conceptually irrelevant video shown before and after explanation. Results showed that, although the use of multimedia can be well-intentioned, it might actually have negative effects on student understanding of scientific concepts because students continually performed better in conditions with less multimedia input.

In a subsequent study, Mayer, Hagerty, Mayer, and Campbell (2005) conducted an investigation to further test the benefits of static images vs. animations. As in the previous study, the authors conducted four experiments with the intent to determine: (1) if static images were more beneficial than animated narrations, (2) if static images were less beneficial than animations, or (3) if static and animated images were equally beneficial. The authors argued that previous literature investigating this topic did not provide clear results due to methodological flaws. Therefore, their four experiments involved college students to compare test scores in retention and transfer of knowledge when learning occurred via computer animation or via text on with static images. The four topics participants learned about were (1) how lightning is created, (2) how a toilet functions, (3) how ocean waves are created, and (4) how a car’s brake system works. Results showed that students participating in the computer animation condition did not outscore the text on paper with static images group in any of the experiments. In addition,
students who learned via text on paper with static images scored significantly higher for retention on two of the experiments. On two of the other experiments, students learning via paper on text with static image significantly outscored students learning via computer animation on the transfer of knowledge tests. Again, these results lead Mayer to state that animation does not appear to be more helpful than static images when learning science content.

Mayer and Moreno (2003) gave suggestions to decrease the cognitive load that students may experience when participating in multimedia learning. They offer strategies to handle situations when combining auditory input with visual input result in cognitive overloading. Although the current study does not investigate animation, it is important to note that these studies show that combining static images with text can be beneficial, within certain limits to the cognitive load.

**Summary**

The research reviewed in this chapter highlighted the fact that very few studies have investigated the thinking process of emerging bilinguals on academic tasks. The few studies that exist do not use a holistic view of bilingualism, and their results may not be valid representations of how linguistically diverse students solve problems. This lack of research is problematic for the assessment community, which has little evidence to create validity arguments to justify the use of tests with linguistically diverse populations. Test developers have no empirical data about the thinking processes of emerging bilinguals solving test items to inform their creating of items. Although verbal protocols have been used with monolinguals to investigate if the thinking processes item writers intend to elicit from students are observed, research has yet to utilize this tool with emerging bilinguals. Rather, current practice relies on testing accommodations to minimize the role of English proficiency in emerging bilinguals’ performance. Unfortunately, there is no evidence that these accommodations reduce the role of language in student
performance. More research is needed to examine the detailed thinking processes of emerging bilinguals solving test items with and without a given accommodation.
Chapter 3
Methodology

Strategy of Inquiry

This sequential mixed methods study (Creswell, 2009) utilizes qualitative techniques to investigate a) how students make sense of items, b) what strategies students use to solve the items, and c) how students use vignette illustrations when solving multiple-choice science items. The data is analyzed to identify emerging codes that explore the experiences of emerging bilinguals and mainstream students when making sense of the items and employing a specific strategy to solve the items. A parallel coding scheme is used to identify the ways emerging bilinguals and mainstream students use illustrations when solving items. After qualitatively analyzing the data, quantitative analyses address the trends across items utilizing frequency counts, tests of statistical significance of frequency differences, and tests of statistical significance of student performance.

By utilizing mixed methods, I am able to provide insight at a detailed level as well at a broader level, across the entire data set. Qualitative analysis allows me to investigate the data in great detail, examining how students interact with items. This detailed analysis allows me to examine the heterogeneity within and across comparative groups. Quantitative analysis allows me to examine the magnitude of similarities and differences across comparative groups.

Setting and participants

This study includes 36 native Spanish-speaking students classified as ELL and 36 native English-speaking students in Grades 6-8 from an English-only middle school in a Western Mountain state. Table 1 summarizes the demographic characteristics of all participants.
### Table 1

*Demographics of Emerging Bilinguals and Mainstream Students*

<table>
<thead>
<tr>
<th></th>
<th>ELLs (n = 36)</th>
<th>Mainstream (n = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td><strong>Average CSAP Score, Grade 6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>445</td>
<td>592</td>
</tr>
<tr>
<td>Reading</td>
<td>567</td>
<td>674</td>
</tr>
<tr>
<td>Writing</td>
<td>474</td>
<td>555</td>
</tr>
<tr>
<td>Science</td>
<td>364</td>
<td>553</td>
</tr>
<tr>
<td><strong>Average CSAP Score, Grade 7</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>483</td>
<td>630</td>
</tr>
<tr>
<td>Reading</td>
<td>578</td>
<td>729</td>
</tr>
<tr>
<td>Writing</td>
<td>485</td>
<td>600</td>
</tr>
<tr>
<td><strong>Average CSAP Score, Grade 8</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>455</td>
<td>601</td>
</tr>
<tr>
<td>Reading</td>
<td>544</td>
<td>659</td>
</tr>
<tr>
<td>Writing</td>
<td>472</td>
<td>546</td>
</tr>
<tr>
<td><strong>Average CELA Score, Grade 6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>569</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average CELA Score, Grade 7</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>556</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average CELA Score, Grade 8</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>565</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of students classified FEP</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Number of students classified LEP</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Number of students classified NEP</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Average Number of Years in U.S.</strong></td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

1CSAP scores are from the spring of the previous academic year. Because the science portion of the CSAP is only administered in Grades 5 and 8, no science scores are available for Grade 7 and Grade 8.

2CELA scores reported are from the fall of the current school year and consist of the composite score of reading, writing, listening, and speaking. The cut-off to be considered fully English proficient is 573 for Grade 6, 574 for Grade 7, and 575 for Grade 8. CELA scores are only administered to non-native English speakers. Therefore, no scores are shown in the table for mainstream students.

3The average number of years in the U.S. is calculated for those students who reported not being born in the country, and is therefore not given for native English speakers as they all reported being born in the U.S.
Similar to the entire population of ELLs, the participants in this study came from a variety of backgrounds, had lived in the country for varying amounts of time, and were classified at different levels of English proficiency. As the table above shows, the average CSAP score for ELLs was lower than the average score for non-ELLs, at every grade level. This score difference is not uncommon, as mainstream students typically outscore ELLs on standardized tests (see Kieffer, Lexaux, Rivera, & Francis, 2009).

Twenty-four ELLs were classified as Limited English Proficient (LEP), while nine students were classified as Fully English Proficient (FEP), and three students were classified as Non-English Proficient (NEP). Note, all classifications were provided by the school and are based on the state assessment for English proficiency (CELA), teacher judgment, and/or parent consent. Fourteen ELLs reported being born in the U.S., while the majority, 20 students, reported that they were born in Mexico. Of the remaining two students, one was born in Puerto Rico and one was born in El Salvador. For those students not born in the country, the average years they had lived in the country was seven years. However, the range of years living in the country spanned from 1-13 years. Due to the heterogeneity of these demographic characteristics, I consider each bilingual to have his or her own unique linguistic repertoire that is different from all other bilinguals. This sample of participants most likely reflects the entire sample of Spanish speaking ELLs at this school as over 80% of students classified as ELL agreed to participate in the study.

Although all mainstream students reported that they were born in the U.S., it is likely that, similar to emerging bilinguals, they had different experiences learning English. Regardless of the fact that each monolingual varies in English proficiency, all mainstream participants reported speaking only English and had never been classified as ELL.
**Researcher Role**

First, it is essential to acknowledge my role as a research assistant for the ITELL project that I have participated in since its inception. The main motivation for this study was my interest in providing empirical evidence to examine how emerging bilinguals and mainstream students make sense of and solve multiple-choice items. While completing work for a course at the university, I became aware of the lack of research conducted on problem solving strategies among bilingual individuals. I was immediately interested in how my research could contribute to filling this gap. Simultaneously, my work with the ITELL project began to provide an opportunity to investigate the problem solving strategies of ELLs when taking multiple-choice items with illustrations as a new form of testing accommodation. These events, combined with my prior experience teaching emerging bilinguals, and desire to improve their education, motivated me to pursue this study. Therefore, while this dissertation is part of the broader ITELL study, the data I collected and analyzed are specific to this dissertation.

My participation in the ITELL project required interaction with the participants’ science teachers and English as a Second Language Teacher. Through these interactions, I gained insight, and at times insider information, about various aspects of daily life at the school in which the study was conducted. This information helped me to familiarize myself with the research site. Overall, this contact with teachers led me to hold a positive opinion about the school environment.

Finally, I am a bilingual, middle class white female. I recognize that, although I speak the same two languages that the participants in this study speak, my background is very different from theirs. I strive to remain cognizant of this difference and let the data speak for itself, not allowing my background to influence interpretation.
**Item Selection**

In order to create a pool of ten Grade 8 multiple-choice science items for this study, I first created a pool of over 20 released items of various assessment programs, including: CSAP (Colorado State Assessment Program), TIMSS (Trends in International Mathematics and Science Study), CST (California Standards Test), and AIMS (Arizona Instrument to Measure Standards). States utilize these assessments to test the overall science knowledge of their student population.

Next, from this pool, I selected items to reflect the proportions of items that assess each of the five standards in the state of Colorado as assessed by the Grade 8 CSAP. The purpose of this state mandated test is to provide “a once yearly snapshot of student progress relative to the Colorado Model Content Standards” (Colorado Department of Education, 2007). Due to the similarity of Standard 1 (Scientific Processes) and Standard 5 (Nature of Understanding Science), these standards were treated as one standard. The pool of over 20 items was narrowed to 10 items by (a) identifying items that could be illustrated according to the ITEL protocol, and (b) identifying linguistic aspects that could be potentially challenging to emerging bilinguals based on characteristics previously identified in the research, such as cloze questions and difficulty vocabulary (Abedi, Hofstetter, & Lord, 2004; Kopriva, 2007). illustrates the items used in this study and their corresponding standards. The table also shows the number and nickname that I assigned to each item for use in this study.

At this school, Standard 2 is taught in Grade 6, Standard 3 is taught in Grade 7, and Standard 4 is taught in Grade 8. Standards 1 and 5 are taught in all grade levels. Because different standards are taught in different levels, younger students may not have had the opportunity to learn about specific content areas. This aspect is considered in the qualitative analysis when
Table 2

*Items Used in the Investigation*

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Item Nickname</th>
<th>Standard</th>
<th>Item Stem</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Seedling</td>
<td>Life science</td>
<td>What is the primary function of the large leaves found on seedlings growing in the forest?</td>
<td>TIMSS, 1999</td>
</tr>
<tr>
<td>2</td>
<td>Moon</td>
<td>Earth and space science</td>
<td>The Moon produces no light and yet it shines at night. Why is this?</td>
<td>TIMSS, 1999</td>
</tr>
<tr>
<td>3</td>
<td>Alexander Fleming</td>
<td>Scientific processes</td>
<td>Alexander Fleming noticed that bacteria growing on a plate of agar did not grow next to a mold growing on the same plate. He wrote in his laboratory report: “The mold may be producing a substance that kills bacteria.” This statement is best described as…</td>
<td>TIMSS, 1999</td>
</tr>
<tr>
<td>4</td>
<td>Stopwatch</td>
<td>Scientific processes</td>
<td>Maria wanted to measure the amount of time it took for a ball to roll down a ramp. She had never used a stopwatch before. Kevin gave her the following directions, but they were in the wrong order. How should she arrange Kevin’s steps so they are in the correct order?</td>
<td>AIMS, 2009</td>
</tr>
<tr>
<td>5</td>
<td>Helium Balloon</td>
<td>Physical science</td>
<td>A balloon filled with helium gas is set free and starts to move upward. Which of the following best explains why the helium balloon moves upward?</td>
<td>TIMSS, 2003</td>
</tr>
<tr>
<td>6</td>
<td>Pine Wood</td>
<td>Physical science</td>
<td>A piece of pine wood floats on the surface of a lake because the water exerts…</td>
<td>CST, 2009</td>
</tr>
<tr>
<td>7</td>
<td>Magnesium</td>
<td>Physical science</td>
<td>When magnesium (Mg) metal is burned in the presence of oxide (O₂), magnesium oxide (MgO) is produced. The properties of magnesium oxide are different than the individual properties of magnesium and oxygen because magnesium oxide is…</td>
<td>CST, 2009</td>
</tr>
<tr>
<td>8</td>
<td>Platypus</td>
<td>Life science</td>
<td>A small animal called the duckbilled platypus lives in Australia. Which characteristic of this animal shows that it is a mammal?</td>
<td>TIMSS, 1999</td>
</tr>
<tr>
<td>9</td>
<td>Cold Region</td>
<td>Life science</td>
<td>Which statement best explains why mammals are found in very cold regions of the world but lizards are not?</td>
<td>AIMS, 2009</td>
</tr>
<tr>
<td>10</td>
<td>Earth’s Temperature</td>
<td>Earth and space science</td>
<td>If the temperature of Earth rose over time, which of the following would occur?</td>
<td>CSAP, 2002</td>
</tr>
</tbody>
</table>
interpreting students’ reasoning. However, in the quantitative analyses, data are not disaggregated due to the restricted number of students within each grade.

In their original version, these items did not contain any illustrations. One item was not illustrated because it was the item that all students took in the non-illustrated form. The other nine items were illustrated because they were given to the students in both the illustrated and non-illustrated form.

**Illustration Development**

The illustrated versions of the nine items were created utilizing a preliminary version of the illustration development protocol from the broader ITELL project. According to this earlier version of the procedures (see Solano-Flores, 2010a; 2010b), the content of an item’s illustration was determined by science teachers and project staff, based on a set of features allowed and a set features not allowed. Those features allowed included: (a) focusing on a specific object or event stated in the stem of an item, (b) an example from a set of cases, (c) features common to a class of objects, (d) a comparison of objects, (e) the parts that comprise a whole, or (f) the basic components of an object. Those features not allowed included: (a) an object mentioned in an answer choice, (b) multiple sages or actions, (c) objects in different scales, (d) sequence of events or processes, (e) hierarchical relationships between objects, (f) references to experiences unique to only some individuals (e.g., inside jokes), (g) symbols (e.g., arrows), and (h) aspects not visible to the human eye. The goal of these features was to systematically design images, based on the analysis of the linguistic properties of the items’ stems and the potential linguistic challenges these properties may pose to ELLs.
Verbal Protocol Development

A verbal protocol was created to probe students’ comprehension of items, thinking processes, and use of images while they solved items with and without illustrations. Over a year period, several iterations of the verbal protocol were developed. I made modifications based on information gained from reviewing the literature, piloting the protocol, input from teachers participating in the ITELL project, and comments from a member of the ITELL technical advisory board with expertise in cognitive interviews. These modifications included: (1) adding information to the introduction to explain that researchers were interested in students’ reasoning when solving science items, (2) informing students that their participation was not for a grade, and (3) placing all follow-up questions after the concurrent and retrospective reporting with the goal of not influencing student reasoning during the think aloud.

Once the protocol was developed in English, I translated it into Spanish. Upon completion of my translation, a native Spanish-speaker from Mexico and member of the ITELL research team reviewed my translation and made changes. These changes ensured that the language use in the protocol more closely followed a Mexican dialect of Spanish, as the majority of the Spanish speaking students had a Mexican origin or descent. The final versions of the protocol were used as a guide to maintain consistency across all student interviews. See Appendix A for the English version of the protocol and Appendix B for the Spanish version.

Interview Procedures

Students participated in the interviews individually. Upon completing the student assent form, I began recording our interactions. The recording of initial interactions provided information about language use during exchanges that were not considered part of the actual interview. The recording continued throughout the students’ verbalizations in concurrent
reporting, retrospective reporting, and follow-up questions. On average, the interview lasted twenty minutes.

During the interview, each student received four items. Two items were illustrated and two items were not illustrated. One of the illustrated items and one of the non-illustrated items were the same for all participants. That is, all students solved the same illustrated item (Item 1) and the same non-illustrated item (Item 2). The remaining two items (one illustrated item and one non-illustrated item) were randomly assigned from the remaining pool of eight items (Items 3-10).

Think-aloud sessions, or verbal protocols, were the main form of data collection. The protocol includes five phases: introduction, warm-up, concurrent reporting, retrospective reporting, and follow-up probing questions. During the introduction phase, I introduced myself and the project to the student. I explained that the purpose of the study was to investigate what students think when they solve science multiple-choice items. I did not mention the illustration at this time, as I my intent was to capture students’ thinking processes as they would occur in a natural testing environment. I also let bilingual students know they could use either language at anytime during the interview. During the warm-up section of the interview, students participated in solving two simple mathematics items to familiarize themselves with concurrent reporting, retrospective reporting, and answering probing question sections.

Once the warm-up was completed, the students participated in concurrent reporting which required them to verbalize their thoughts as they completed the task. Students were given the item and asked to read it aloud from beginning to end (they did not read the item silently to themselves before reading the item aloud). If a student fell silent, I followed recommendations given by Ericsson and Simon (1993) that instruct the interviewer say only one thing: “keep talking” or “habla.” As previously mentioned, Ericsson and Simon point out that prompts such
as, “tell me what you’re thinking,” can trigger metacognitive processes not naturally occurring in the participant’s mind. The focus of this study is to investigate students’ thinking processes that occur naturally in assessment situations and I attempted to minimize my influencing students’ thinking in anyway. Therefore, only “keep talking” or “habla” was said to students falling silent.

The concurrent reporting was immediately followed by the retrospective reporting for the same item in which students recalled their steps to solve that item. Once all concurrent and retrospective reporting was completed for the four items in the interview, students answered follow-up questions. These questions investigated what students found difficult or confusing, how the illustration was used, and their interpretation of what the item asked them to do.

All interviews were transcribed verbatim to reflect students’ exact verbalization. Bilingual interviews were transcribed by a bilingual transcriptionist to capture the students’ use of both languages. In addition, I reviewed each transcription while listening to the audio file to ensure accuracy.

For the purposes of this investigation, the data analysis focused on three sections of the interview: concurrent reporting, retrospective reporting, and follow-up questions. Because the introductory and warm-up sections of the interview were designed to familiarize the participants with the procedures of the interview, these sections were not of interest for the main analysis.

**Language use during the interview.** I began interviews by re-informing bilinguals that I spoke both English and Spanish. I also explicitly told bilinguals that they could use any language they preferred at anytime during the interview. Because students attended an English-only school, they have not been accustomed to communicating about academic topics in Spanish. Therefore, I recognize that the location of the interview influenced how students chose to use
language. However, I continually incorporated both English and Spanish throughout the
interviews to create an inviting environment to use both languages at all times.

**Rapport building.** Before beginning the interviews, I spent several weeks in the English as a
Second Language classroom to introduce myself to the bilingual students and build a rapport
with them. During these initial interactions, I made a conscious effort to use English and Spanish
so that they viewed me as a fellow bilingual. I also tried to make it clear that I was not a teacher
or an authoritative figure from the school. With these goals in mind, I hoped to establish a
comfortable and casual relationship with students.

**Coding System**

I developed a coding system to identify all actions students carried out to make sense of
items. The coding system was developed based on an exhaustive analysis of the interview
transcriptions with both bilinguals and monolinguals. Utilizing constant comparative method
(Glaser, 1965), I incorporated aspects identified as potentially challenging to students in
academic texts. These aspects included the testing register (Abedi, 2006; Abedi, Hofstetter, &
Lord, 2004; Solano-Flores, 2006; 2008a) and science vocabulary (Osborne & Wellington, 2001).
Thus, this coding system targeted challenges students may face when reading: (1) terms specific
to the domain of science, (2) terms not specific to the domain of science, (3) terms with more
than one meaning, and (4) terms that are common to scientific discourse.

The coding system captured aspects at macro and micro levels of the ways students made
sense of the items. The first two categories captured students’ actions at a macro level,
identifying (1) challenges students face to understand the item and (2) strategies they use to
understand the item. The last two categories focused on students’ actions at a micro level,
investigating (3) if students deviate from the printed text when reading the item aloud, and (4)
any corrections of these deviations. Respecting sociolinguistic principles, I distinguished between aspects of student interactions with science and non-science aspects of the items. This coding system also captured when bilinguals drew from their native language by incorporating a parallel coding system to capture when any of these actions are carried out using Spanish. Table 3 shows the coding system for how students made sense of items. It provides the code, category, definition, example, and example location for all codes. The first part of the coding system, Overall Comprehension, tracked students’ general understanding of the item. For example, students could have forgotten an aspect of the item after reading it in entirety. Or, students could report that an aspect of the item was confusing. Finally, I identified when students did not understand science terms, non-science terms, or when they understood individual words but could not make sense of the item at a syntactical level.

The second part of the coding system, Strategies to Comprehend, identified specific actions students carried out to better make sense of the item. Here, it was possible for bilingual students to draw from their native language to carry out a specific strategy, such as self-monitoring their comprehension. Or, a student could restate the problem in their own words, using English, Spanish, or a combination of both. In addition, students could have drawn from the use of cognates or need to reread the question to make sense of it.

The section, Read Aloud Deviations, utilized a bilingual view of the running record analysis to investigate students’ interactions with the item at a micro level. Focusing on the strategic behavior students carried out when they read aloud, this part of the coding system identified when students’ verbalizations deviated from the printed text. Note, while the focus was deviations, these actions have been shown to be strategic behavior of good readers (Clay, 2002). Deviations were not considered incorrect, but rather different from the printed text.
### Challenges to Understanding

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>OU1</td>
<td>As is</td>
<td>Student (S) reads item as printed</td>
<td>So, uhh I think it’s uhh, what is it? So I think it’s B. Wait, what’s the question?</td>
<td>T. 1-27-47, RA: line 256</td>
</tr>
<tr>
<td>OU5</td>
<td>Forgets</td>
<td>S unable to recall item content after reading item</td>
<td>Why does it confuse you? Can you explain why? Um it’s cuz at first I read the moon produces no light and then I read and yet it shines at night so it says it shines but then it shines.</td>
<td>T. 1-22-37, FU: line 261</td>
</tr>
<tr>
<td>OU6</td>
<td>Confused</td>
<td>S is confused by an aspect of the item</td>
<td>Why does it confuse you? Can you explain why? Um it’s cuz at first I read the moon produces no light and then I read and yet it shines at night so it says it shines but then it shines.</td>
<td>T. 1-22-37, FU: line 261</td>
</tr>
<tr>
<td>OU7</td>
<td>Science term</td>
<td>S reports not knowing the meaning of a science term</td>
<td>A piece of pine wood floating floats on the surface of the lake because the water ex ex ex tracts, extends, I don’t know what that word is.</td>
<td>T. 1-27-47, RA: line 178</td>
</tr>
<tr>
<td>OU8</td>
<td>Non-science term</td>
<td>S reports not knowing the meaning of a non-science term</td>
<td>Why does it confuse you? Can you explain why? Um it’s cuz at first I read the moon produces no light and then I read and yet it shines at night so it says it shines but then it shines.</td>
<td>T. 1-22-37, FU: line 261</td>
</tr>
<tr>
<td>OU9</td>
<td>Difficulty at syntactical level</td>
<td>S reports not understanding part of the item at the phrase or sentence level</td>
<td>So looking at the question, were there specific words in there that were hard for you? “No, but I didn’t got it that much so,” No? And so do you know what specific part you weren’t getting? “All of it.”</td>
<td>T. 1-27-47, RA: line 178</td>
</tr>
</tbody>
</table>

### Strategies to Comprehend

<table>
<thead>
<tr>
<th>Code</th>
<th>Monitor</th>
<th>Language</th>
<th>Description</th>
<th>Example</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC1</td>
<td>English</td>
<td>S tracks h. comprehension of the item in English</td>
<td>The moon produces no light, and yet it shines at night. Why is this? First I’m gonna read it again, ’cause it didn’t really get it.</td>
<td>T. 0-52-06, RA: line 113</td>
<td></td>
</tr>
<tr>
<td>SC2</td>
<td>Spanish</td>
<td>S tracks h. comprehension of the item in Spanish</td>
<td>What is the primary function of large leaves found on the seedlings growing in the forest? OK, so why are there big leaves on small trees?</td>
<td>T. 0-52-06, RA: Line 166</td>
<td></td>
</tr>
<tr>
<td>SC5</td>
<td>English</td>
<td>S restates the stem into own words in English</td>
<td>C porque decia que sea level would rise y se que es real porque se que el oceanno esta subiendo y que tambien dice que polar ice caps would decrease in size que decrease signifcara como subiendo supe, eso si como decrease como, I think, si como, bajando, si, eso.</td>
<td>T. 1-24-42, RT: line 152</td>
<td></td>
</tr>
<tr>
<td>SC6</td>
<td>Spanish</td>
<td>S restates the stem into own words in Spanish</td>
<td>What is the primary function of large leaves found on the seedlings growing in the forest? OK, so why are there big leaves on small trees?</td>
<td>T. 0-52-06, RA: Line 166</td>
<td></td>
</tr>
<tr>
<td>SC7</td>
<td>Translate</td>
<td>S translates aspect of stem to Spanish, or uses cognates</td>
<td>“Un poco dificil. Y que tal estas, polar ice caps?” Esto lo se, como capas del hielo.</td>
<td>T. 1-10-18b, FU: line 49</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>---------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>SC9</td>
<td>Reread</td>
<td>S reads the item again</td>
<td><em>I read, first I read the questions and the answers and I tried I read the question again, cuz I didn’t get it that well and then I went to look at the answers again and I just knew it was A.</em></td>
<td>T. 1-22-37 RT: line 155</td>
<td></td>
</tr>
<tr>
<td><strong>Read Aloud Deviations</strong></td>
<td>Repeat</td>
<td>S repeats word while reading aloud</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA1</td>
<td>Science term</td>
<td></td>
<td><em>How should she arrange Kevin’s steps so they are in the correct order?</em></td>
<td>T. 1-36-65 RA: line 161</td>
<td></td>
</tr>
<tr>
<td>RA2</td>
<td>Non-Science term</td>
<td></td>
<td><em>Which characteristic of this animal shows that it is a mammal? It eats other animals. It feeds its young milk. It makes a nest and lays eggs. It has webbed feet.</em></td>
<td>T. 0-71-56 RA: line 129</td>
<td></td>
</tr>
<tr>
<td>RA2b</td>
<td>Phrase with both science and non-science terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA3</td>
<td>Sound out</td>
<td>S sounds out syllables of word while reading aloud</td>
<td><em>the moon has many cra-cra-ters</em></td>
<td>T. 1-13-23 RA: line 164</td>
<td></td>
</tr>
<tr>
<td>RA4</td>
<td>Non-Science term</td>
<td></td>
<td><em>To get rid of ex-exssssess water that is entering through the roots.</em></td>
<td>T. 1-32-54 RA: line 148</td>
<td></td>
</tr>
<tr>
<td>RA5</td>
<td>Produce Non-Word</td>
<td>S speaks word that is not officially a word</td>
<td><em>downward force equal to the weight of the dis-ment water</em></td>
<td>Transcript 1-27-47 RA: line 189</td>
<td></td>
</tr>
<tr>
<td>RA6</td>
<td>Non-Science term</td>
<td></td>
<td><em>A, to proveel shade for the root system,</em></td>
<td>T. 1-04-08 RA: line 126</td>
<td></td>
</tr>
<tr>
<td>RA7</td>
<td>Omit</td>
<td>S leaves out a word when reading aloud</td>
<td><em>to allow the leaf damage by insects, to gather as much light as possible for.</em></td>
<td>T. 1-01-01 RA: line 120</td>
<td></td>
</tr>
<tr>
<td>RA8</td>
<td>Non-Science term</td>
<td></td>
<td><em>C the moon covered with a thin layer of ice.</em></td>
<td>T 1-06-11 RA: line 146</td>
<td></td>
</tr>
<tr>
<td>RA9</td>
<td>Mispronounce</td>
<td>S pronounces a word different from the standard form</td>
<td><em>To gather as much light as possible for foto-SIGN-thesis</em></td>
<td>T. 1-32-54 RA: line 150</td>
<td></td>
</tr>
<tr>
<td>RA12</td>
<td>Non-Science term</td>
<td></td>
<td><em>to a-low for leaf damage by insects</em></td>
<td>T. 1-12-21 RA: line 169</td>
<td></td>
</tr>
<tr>
<td>RA15</td>
<td>Insert</td>
<td>S adds a word to those written in the item</td>
<td><em>What is the primary function of large forest leaves found on the seedlings growing in the forest?</em></td>
<td>T. 0-52-06 RA: line 164</td>
<td></td>
</tr>
<tr>
<td>RA16</td>
<td>Non-Science term</td>
<td></td>
<td><em>Hold up</em> the watch in your hand</td>
<td>T. 1-27-47 RA: line 215</td>
<td></td>
</tr>
<tr>
<td>SB1</td>
<td>Substitutions</td>
<td>S substitutes a word for those written on the page</td>
<td><em>B, to get rid of existing water that is entering through the roots</em></td>
<td>T. 1-04-08 RA: line 126</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------</td>
<td>------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>SB2</td>
<td>English: meaning kept</td>
<td>What is the primary fraction of the large leaves found on the seedling growing in the forest?</td>
<td>T. 1-27-47</td>
<td>RA: line 251</td>
<td></td>
</tr>
<tr>
<td>SB4</td>
<td>Spanish: meaning changed</td>
<td>No example found</td>
<td>T. 1-12-21</td>
<td>RA: line 168</td>
<td></td>
</tr>
<tr>
<td>SB5</td>
<td>Spanish: meaning kept</td>
<td>No example found</td>
<td>T. 1-12-21</td>
<td>RA: line 185</td>
<td></td>
</tr>
<tr>
<td>SB7</td>
<td>English: meaning changed</td>
<td>To divide shade for the root system</td>
<td>T. 1-12-21</td>
<td>RA: line 146</td>
<td></td>
</tr>
<tr>
<td>SB8</td>
<td>English: meaning kept</td>
<td>The moon is covered with tiny layer of ice</td>
<td>T. 1-12-21</td>
<td>RA: line 185</td>
<td></td>
</tr>
<tr>
<td>SB10</td>
<td>Spanish: meaning changed</td>
<td>Step one: hold the stopwatch in the hand.</td>
<td>T. 1-12-21</td>
<td>RA: line 146</td>
<td></td>
</tr>
<tr>
<td>SB11</td>
<td>Spanish: meaning kept</td>
<td>No example found</td>
<td>T. 1-12-21</td>
<td>RA: line 146</td>
<td></td>
</tr>
<tr>
<td>CU1</td>
<td>Non-word</td>
<td>S properly says a word initially produced as non-word</td>
<td>OK, when ma, when Ma..Magnilium - MG metal is burned, in the presence of Oxi-gent - O2 magnesium oxide MGO is produced.</td>
<td>T. 1-05-10</td>
<td></td>
</tr>
<tr>
<td>CU2</td>
<td>Non-Science term</td>
<td>S properly pronounces after reading it twice</td>
<td>If the temperature of earse-earth rose over time, which of the following would occur?</td>
<td>T. 2-79-67</td>
<td></td>
</tr>
<tr>
<td>CU3</td>
<td>Science term</td>
<td>How should she arrange Kevin’s steps so they are in the order, in the correct order?</td>
<td>T. 1-12-21</td>
<td>RA: line 149</td>
<td></td>
</tr>
<tr>
<td>CU4</td>
<td>Non-Science term</td>
<td>Step one: hold the stop-watch in on in one hand.</td>
<td>T. 1-12-21</td>
<td>RA: line 149</td>
<td></td>
</tr>
<tr>
<td>CU5</td>
<td>Science term</td>
<td>How should she arrange Kevin’s steps so they are in the order, in the correct order?</td>
<td>T. 1-12-21</td>
<td>RA: line 149</td>
<td></td>
</tr>
<tr>
<td>CU6</td>
<td>Non-Science term</td>
<td>Step one: hold the stop-watch in on in one hand.</td>
<td>T. 1-12-21</td>
<td>RA: line 149</td>
<td></td>
</tr>
<tr>
<td>CU7</td>
<td>Science term</td>
<td>A piece of pine wood floats on the surface of a lake because the water exerts: an upward force equals, equal to the weight of the wood.</td>
<td>T. 1-06-11</td>
<td>RA: line 163</td>
<td></td>
</tr>
<tr>
<td>CU8</td>
<td>Non-Science term</td>
<td>Alexander Flemming noticed that bacteria growing on a plate agar did not grow next to the mold that it, that was growing on the same plate.</td>
<td>T. 1-06-11</td>
<td>RA: line 178</td>
<td></td>
</tr>
<tr>
<td>CU9</td>
<td>Science term</td>
<td>To provide shade for the root system, to get rid of exess water to that, that is entering through roots</td>
<td>T. 1-02-03</td>
<td>RA: line 124</td>
<td></td>
</tr>
</tbody>
</table>
For example, students could substitute words when reading aloud, informing these substitutions based on their knowledge of their language system(s). When students substituted words, they informed their choices with visual, syntactical, or semantic information from the other words on the page (Clay, 2002). Substitutions may or may not have changed the meaning of the item. Take the following excerpt from Item 1, Seedling: “What is the primary fraction of the large leaves found on seedlings growing in a forest?” Here, the student changed the meaning of the science term *primary function* by substituting it with *primary fraction*. Most likely, this student was relying on visual and semantic information for this substitution, seeing the word *primary* followed by another word that begins with the letter f which triggered him to read a common mathematical term, *primary fraction*. Students could also substitute words that preserved the printed text’s meaning. An example from Item 2, the Moon: “The moon is covered with a tiny layer of ice” was an example of a student substituting a non-science term (*tiny* in place of *thin*) and maintaining the original meaning. This substitution was most likely informed by both visual and semantic information as both words started with the letters “ti” and referred to a small amount.

Finally, the last section of the coding system, Corrections to Read Aloud, identified when students corrected any action that was identified in the Read Aloud Deviation category. These corrections described when a student changed their original verbalization to match the printed text of the item. They were considered metacognitive activity when reading aloud.

**Focus group.** As part of the process of developing the coding system, I held a focus group to review the coding system in a previous version and improve it based on participants’ knowledge and experiences. The focus group included three bilingual education experts, all holding
doctorates in the field of bilingual education. In addition, they all had experience teaching emerging bilinguals as well as conducting research to investigate aspects of biliteracy.

The focus group’s recommendations resulted in changes to the Read Aloud Deviations and the Corrections to Read Aloud coding categories, including the incorporation of the running record into the Read Aloud Deviation section. The focus group emphasized the need to utilize a bilingual view while using the running record. These changes to the coding system allowed for a more systematic and detailed analysis of students’ experiences making sense of the items.

**Classifying science vocabulary.** To classify terms in the items as science or non-science terms, a professor of science education, who was also a licensed science teacher, categorized items based on the five criteria shown in Table 4.

**Table 4**

<table>
<thead>
<tr>
<th>Code</th>
<th>Type of term</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Terms specific to the domain of science; only used when discussing science</td>
<td>electron, refraction</td>
</tr>
<tr>
<td>2</td>
<td>Terms used in science but also common to everyday conversations</td>
<td>energy, weight</td>
</tr>
<tr>
<td>3</td>
<td>Terms with only one meaning</td>
<td>gain, optimum</td>
</tr>
<tr>
<td>4</td>
<td>Terms with more than one meaning</td>
<td>negative, light, mass</td>
</tr>
<tr>
<td>5</td>
<td>Terms that appear frequently in the scientific discourse</td>
<td>likely, in contrast</td>
</tr>
</tbody>
</table>

The criteria shown in Table 4 were not mutually exclusive and a term could be classified as more than one code. For example, the word *photosynthesis* was a term that was specific to the domain of science (code 1) and was a term with only one meaning (code 3). The professor of science education was given directions to code both the stem and the options for each item used in the
study. See Appendix C for the directions provided to this science expert to classify science terms.

Next, to identify the strategies they used to solve each item, I used constant comparative analysis (Glaser, 1965) by incorporating problem solving strategies identified by previous research (Ericsson & Simon, 1993; Hamilton, Nussbaum, & Snow, 1993) and performing an exhaustive analysis on the transcripts to complete the coding system. Table 5 shows the coding system that captured the strategies students used to answer each item. This section of the coding focused on three main categories: Background Knowledge, In the Moment Knowledge, and Testing Strategies. It was possible for bilingual participants to carry out any strategy using English only, or incorporating Spanish. Therefore, Column 2 provides a code to capture when Spanish is utilized, indicated by the original code followed by an S.

The first category in Table 5, Background Knowledge, refers to instances when students draw from previous experiences to inform their answer choice. This included cases when students drew from their experiences inside the classroom as well as outside the classroom. In addition, students may apply scientific rules or concepts to the item’s content in order to arrive at the answer. The next category, In the Moment Knowledge, applies to cases when students use information from the item to inform their problem solving. For example, students might focus on a specific aspect of the item to find the answer or they might relate the information given in the item to the real world. Finally, the last category, Testing Strategies, lists various tactics students use to solve multiple-choice items. These strategies include, making a guess, using process of elimination, and self-monitoring one’s progress when solving the item.
Table 5

**Coding System: Problem Solving Strategies**

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>BK1</td>
<td>BK1S</td>
<td>Recall facts</td>
<td><em>I think it’s D, “to gather as much light as possible,” because I know that plants need photosynthesis to survive.</em></td>
<td>T. 0-87-80 RT: line 149</td>
</tr>
<tr>
<td>BK2</td>
<td>BK2S</td>
<td>Incorrect recall of facts</td>
<td><em>It’s [the moon] covered with a thin layer of ice so that means that moon shines a lot and so does the ice.</em></td>
<td>T. 1-28-48</td>
</tr>
<tr>
<td>BK3</td>
<td>BK3S</td>
<td>General background</td>
<td><em>And I think it’s C, just because, like, I know I at one time studied this, but I kind of forget.</em></td>
<td>T. 0-73-58</td>
</tr>
<tr>
<td>BK4</td>
<td>BK4S</td>
<td>Classroom experience</td>
<td><em>cause we just studied this in class and Mr. F. taught us that, that’s why they use helium in blimps, because the density is less than air</em></td>
<td>T. 0-52-06</td>
</tr>
<tr>
<td>BK5</td>
<td>BK5S</td>
<td>Non-classroom</td>
<td><em>Um, definitely background knowledge on this, because my dad and I were talking about this, like, last weekend. Um, yeah.</em></td>
<td>T. 0-60-26</td>
</tr>
<tr>
<td>BK7</td>
<td>BK7S</td>
<td>Applies scientific fact</td>
<td><em>So in a growing forest there would be less light, and it would need more light, and mostly seedlings do start out with big leaves to help them get the light they need, so it’s D.</em></td>
<td>T. 0-73-58</td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>BK9</td>
<td>BK9S Refer to related concept</td>
<td>S makes an association between the specific example mentioned in the stem and a more general science concept</td>
<td><em>Well first, since like when you pour water on a seed, you have to wait a few days before it grows root on the bottom and then the plant or tree will grow from the roots.</em></td>
<td>T. 1-18-31 RT: 152</td>
</tr>
<tr>
<td>BK10</td>
<td>BK10S Partial truths</td>
<td>S recalls aspects of the item that are not entirely true</td>
<td><em>Um, a mixture, I think, because it’s two different things and it comes to one.</em></td>
<td>T. 1-34-59 RA: line 114</td>
</tr>
<tr>
<td></td>
<td>In The moment knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK1</td>
<td>MK1S Projection</td>
<td>S answers the question according to how they would carry out the action specified in the item</td>
<td><em>Cuz like I think, cuz that’s the way I would do it. I would do it in those steps, and then um yah I just woulda said it in that order.</em></td>
<td>T. 1-27-47 RA: line 220</td>
</tr>
<tr>
<td>MK3</td>
<td>MK3S Think of real object</td>
<td>S refers to the “real life” version of an aspect mentioned in the stem</td>
<td><em>I kinda like I thought of the real moon like what happens to it so I thought it rotated all over the world so I said question B</em></td>
<td>T. 1-05-10 FU: line 332</td>
</tr>
<tr>
<td>MK4</td>
<td>MK4S Identify helpful information</td>
<td>S points out specific terms or phrases that lead h. to choose an option</td>
<td><em>Well, the first part, he noticed bacteria growing on the plate, and that was the part you need to know, and then the statement, not really, he didn’t do any tests or anything, so yeah.</em></td>
<td>T. 0-87-80 RT: line 136</td>
</tr>
<tr>
<td>MK7</td>
<td>MK7S Visualize</td>
<td>S reports creating a mental image of the item</td>
<td><em>Well, I kind of imagined a plant with big, kind of draping leaves, so that really helped me imagine the water coming down from the leaves and then back to the earth.</em></td>
<td>T. 0-64-38 FU: line 313</td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>MK9</td>
<td>MK9S Incorrectly confirm fact</td>
<td>S states that information given in the item is factually incorrect</td>
<td>I read it, went through all the options, and I got rid of some because—especially, like, C, because I knew it’s not right. And then I knew—and using, like, narrowing down, I could figure out D was the right one.</td>
<td>T. 0-73-58 RT: line 123</td>
</tr>
<tr>
<td>SS1</td>
<td>SS1S Process of elimination</td>
<td>S removes potential answers to narrow choices</td>
<td>I read the question and I went through each one to see if it made sense using reasoning.</td>
<td>T. 0-73-58 RT: line 109</td>
</tr>
<tr>
<td>SS2</td>
<td>SS2S Makes a guess</td>
<td>S chooses an answer based on guessing</td>
<td>I think it’s B, I don’t know why. I’m just gonna guess.</td>
<td>T. 1-19-33 RA: line 210</td>
</tr>
<tr>
<td>SS3</td>
<td>SS3S Makes sense</td>
<td>S recalls selecting an answer that made sense</td>
<td>I kept looking down here to see if what I was doing matched up with the options, and if it didn’t, I’d go back and say, “Wait, this is not what it was, let me try it again.”</td>
<td>T. 0-61-29 RT: line 182</td>
</tr>
<tr>
<td>SS5</td>
<td>SS5S Self-monitoring</td>
<td>S tracks h. progress when solving the item</td>
<td>So, I would think of it then I would try and find the answer again.</td>
<td>T. 1-06-11 RA: line 149</td>
</tr>
<tr>
<td>SS6</td>
<td>SS6S Planning</td>
<td>S devises a course of action that h. will carry out</td>
<td>Not really, just that I don’t really know anything about mammals and lizards.</td>
<td>T. 1-19-33 FU: line 286</td>
</tr>
<tr>
<td>SS7</td>
<td>SS7S Self-evaluation</td>
<td>Make an assessment of his/her ability</td>
<td>I read this really fast to think and I thought it was A, I don’t know why, I just thought it was.</td>
<td>T. 1-05-10 RT: 217</td>
</tr>
<tr>
<td>SS8</td>
<td>SS8S Unable to identify strategy</td>
<td>S is unable to describe how h. solved the item</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The coding systems shown in Tables 3 and 5 allowed me to compare how students make sense of items and the strategies they use to solve items according to ELL status and presence of illustration in an item. The result was a comparison of four groups: (1) ELLs taking items with illustrations, (2) ELLs taking items without illustrations, (3) mainstream students taking items with illustrations, and (4) mainstream students taking items without illustrations.

Finally, I compared how students reported using the illustration when solving the item using the coding system shown in Table 6. If the student provided no information during the concurrent or retrospective reporting about how he/she used the illustration, a follow-up question was used to gather this information. This question asked the students if they (a) saw the illustration, and (b) used it to solve the item. If the student reported that they did use the illustration, I asked them to describe how they used it.

As Table 6 shows, there were five main categories of illustration use and each major category contained 3-4 subcategories. First, students could have reported that they used the illustration to focus on an aspect of the item, the entire item, or an aspect of the illustration. Second, students could have utilized the illustration to make connections to their background knowledge or experiences that occurred in the classroom or outside the classroom. Third, students may have stated that the illustration helps them visualize aspects of the item. Fourth, students may have sought information not given in the text from the illustration, look for an answer, or eliminate some of the options. Finally, some students may not have used the illustration, or they may have found it confusing or interesting.
<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>FO1</td>
<td>Focus</td>
<td>Students use the illustration to…</td>
<td><strong>FO1</strong> Part of the stem center attention on an aspect of the stem</td>
<td>T. 0-59-24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Not that much, but a little, yeah, because I could see the leaves on top and I could tell that it was showing—’cause it was showing this and it was also a visual representation, so I could—I could just—because I'm more visual, I could—it helped represent the seedling and show what the question was talking about in a different way.</em></td>
<td>FU: line 341</td>
</tr>
<tr>
<td>FO2</td>
<td>Entire item</td>
<td>center attention on the whole stem</td>
<td><strong>FO2</strong> Entire item center attention on the whole stem</td>
<td>T. 0-76-63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>I usually just looked at it, just to, like, see, just to keep reminding myself what the question was asking without having to read it over again.</em></td>
<td>FU: line 207</td>
</tr>
<tr>
<td>FO3</td>
<td>Part of illustration not mentioned in stem</td>
<td>center attention on an aspect of illustration that is not mentioned in the stem</td>
<td><strong>FO3</strong> Part of illustration not mentioned in stem center attention on an aspect of illustration that is not mentioned in the stem</td>
<td>T. 0-63-35</td>
</tr>
<tr>
<td></td>
<td>Make connections</td>
<td></td>
<td><em>Hmm. OK, so you can see the ripples. I would say C.</em></td>
<td>RA: line 145</td>
</tr>
<tr>
<td>MC1</td>
<td>Think about scientific knowledge</td>
<td>recall a scientific concept or topic</td>
<td><strong>MC1</strong> Think about scientific knowledge recall a scientific concept or topic</td>
<td>T. 0-63-35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Um, I thought about what I knew about plants and how they make their food from light, and yeah.</em></td>
<td>FU: line 311</td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>MC2</td>
<td>Think about classroom experience</td>
<td>recall an experience from school</td>
<td>When I first looked at the picture, I thought about, um, mostly just, like—last year we looked—we had studied plants last year in fifth grade, and we looked at this, um, cross-cut section of a plant, so as soon as I looked at that, I saw the little thing and I saw the cross-cut of the plant.</td>
<td>T. 0-60-26 FU: line 318</td>
</tr>
<tr>
<td>MC3</td>
<td>Think about outside classroom experience</td>
<td>recall an experience from outside of school</td>
<td>I actually thought of, like, a little seedling in the garden, because my grandma and I used to grow a garden in her backyard. There would be trees around the border, surrounding the field, and we’ve have, like, plants growing in throughout the trees. So, like, using my knowledge of gardening to answer questions sometimes helps.</td>
<td>T. 0-61-29 FU: line 331</td>
</tr>
<tr>
<td>MC4</td>
<td>General background knowledge</td>
<td>recall related knowledge previously known</td>
<td>I thought about how that could be relevant to my background knowledge.</td>
<td>T. 065-39 FU: line 245</td>
</tr>
<tr>
<td>VI1</td>
<td>Part of the stem</td>
<td>envision a specific aspect of the stem</td>
<td>The picture kind of helped, ’cause then I could visualize the large leaves.</td>
<td>T. 0-64-38 FU: line 288</td>
</tr>
<tr>
<td>VI2</td>
<td>Entire stem</td>
<td>envision the entire item</td>
<td>Well it’s talking about how the large leaf found on the forest, so I kind of saw it and pictured the forest and saw the leaves and like tried to like connect it to one of these answers.</td>
<td>T. 1-01-01 FU: line 171</td>
</tr>
<tr>
<td>VI3</td>
<td>Real version of item(s) mentioned in stem</td>
<td>envision an aspect as it appears in the world</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>S11</td>
<td>Look for answer</td>
<td>seek out the correct option</td>
<td><em>The picture, ‘cause—yeah. When I read the question and then got down to B and D, I started to think about it a little more and then the picture helped me, ‘cause it wasn’t showing it raining or anything, and there was no water, and it looked like there was light on it.</em></td>
<td>T. 0-50-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FU: line 255</td>
</tr>
<tr>
<td>S12</td>
<td>Support answer choice</td>
<td>justify why they chose an option</td>
<td><em>Well, yah just that he’s just like in a hole, like in a nest and all that and he was like with eggs and all that. So it reminded you of his nest and his eggs.</em></td>
<td>T. 1-12-21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FU: line 384</td>
</tr>
<tr>
<td>S13</td>
<td>Eliminate wrong answer</td>
<td>remove potential options</td>
<td><em>I cancelled out all the ones that I knew couldn’t be that, and then I looked at the picture to see what was going on in the picture, and I saw that there was light around it, so I crossed off another one, and then I was left with one.</em></td>
<td>T. 0-63-35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FU: line 195</td>
</tr>
<tr>
<td>S14</td>
<td>Give student heads up</td>
<td>predict what an item is about</td>
<td><em>Well, as I said before, it gives you, like, a heads-up of what it’s about. (Do you remember what you thought of when you saw the picture?) I thought, “Oh, it’s probably about temperature.”</em></td>
<td>T. 0-83-74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FU: line 180</td>
</tr>
<tr>
<td>Code</td>
<td>Category</td>
<td>Definition</td>
<td>Example</td>
<td>Location</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>SI5</td>
<td>Understand</td>
<td>better comprehend the text of the item</td>
<td>Some of the words I just didn’t know, and when they said “plate” it wasn’t generally specific and stuff, and then I looked at the picture and I saw what kind of plate and stuff like that.</td>
<td>T. 0-60-26 FU: line 222</td>
</tr>
<tr>
<td>IU1</td>
<td>Did not use</td>
<td>ignore or do not acknowledge using image</td>
<td>Well, it—maybe it was saying something about the temperature, but I don’t think it really fit in the question at all. That was just kind of strange, and it didn’t have any labels, and I didn’t know what that was for, so I just discarded it.</td>
<td>T. 0-59-24 FU: line 266</td>
</tr>
<tr>
<td>IU3</td>
<td>Interesting</td>
<td>describe the image as appealing</td>
<td>I liked the picture being there, but it didn’t really help me solve it, but it’s just, like, it—I don’t know why. It’s just, I like ’em with the pictures.</td>
<td>T. 0-52-06 FU: line 214</td>
</tr>
<tr>
<td>IU4</td>
<td>Confusing</td>
<td>are unclear what the image means</td>
<td>A thermometer, ok I just wanted to know if it looks like a thermometer or if it was a bad drawing. So, again what was confusing about the thermometer? Like I don’t know like how does it go with the question.</td>
<td>T. 1-01-01 FU: line 243</td>
</tr>
</tbody>
</table>
Coding Procedures

The original intent was to train a second coder to independently code transcripts to identify the problem solving strategies and uses of the vignette illustration. Due to the complexity of the coding system, time did not permit independent coding. In addition, there were aspects of the science content knowledge that were beyond my expertise. Moreover, there is evidence that consensus-based coding by individuals with different professional backgrounds is justifiable in order to address the multidimensionality of events related to culture and language (Solano-Flores, Backhoff, & Contreras-Niño, 2009). Consequently, I decided consensus-based coding was the best option for coding the data. These coding procedures allowed the second coder, a former science teacher, and I to discuss each transcript together.

The procedure for consensus coding was as follows. First, all transcripts were randomized and identifying information about ELL status and grade level was removed. The second coder and I held a detailed discussion of the coding system to ensure that we both understood the codes. Next, we separately coded one transcript selected at random. Then, we compared our coding and discussed all disagreements by identifying which portion of the transcript we used as evidence to support the coding choice and the coding system definitions to apply the correct code. This discussion continued until both coders agreed that the appropriate code had been utilized.

Due to time limitations, I was the only coder who coded how students made sense of the item, Part 1 of the coding system. This coding involved listening to the audio recording of the interview while following along with the transcription to make sure each interview transcript accurately followed student verbalizations. Next, I identified the presence of the codes listed in
the coding system, which were typically low-inference codes (Carspecken, 1996) that are not at great risk of being misinterpreted.

**Data Analysis**

**Qualitative analysis.** Upon coding the transcripts using the coding systems described above, I began an ongoing process of continual reflection on the data. For each item, I examined themes in the data by item, linguistic group, and presence of illustrations and wrote a short summary, or memo for each case (Creswell, 2005). In these memos, I drew from student verbalizations in all three sections of the interview (concurrent reporting, retrospective reporting, and follow-up questioning) to provide supporting evidence of my claims. Next, I began to make sense of the data as a whole by clustering similar observations together to draw out larger themes across items. I also noted confirming and disconfirming evidence to support my claims.

**Verification procedures.** As previously mentioned, combining data from student verbalizations in the concurrent reporting, retrospective reporting, and probing questions was a source for triangulating data (Ericsson & Simon, 1993; Taylor & Dionne, 2000). Additionally, I presented disconfirming evidence that illustrated cases that do not follow the overall trend. Maxwell (1996) explains that rigorous examination of supporting and discrepant data is a crucial part of making appropriate conclusions on your data. Therefore, careful attention was paid to those cases that do not fit the overall pattern of the data.

**Quantitative analysis.** Quantitative analysis was based on frequency counts of the codes in the coding system. I examined differences across both linguistic status and presence of vignette illustration. I assumed exchangeability for Items 3-10 and collapsed them into one pool to carry out the frequency counts. Although individual items vary in their linguistic features and content assessed, collapsing the items was necessary due to the small number of students responding to
each item. For this pool, if Student A received Item 4, Stopwatch, and Student B received Item 8, Platypus, they were considered as the same items. This approach allows for the generalization across different item types.

Items 1 and 2 were analyzed separately. Because all students answered these two items, including them in the pool would result in these items receiving greater weight than the other items. Analyzing these two items allows for the generalization across all participants.

Quantitative analysis examined trends at two levels. First, at a macro level, I used the coding clusters or categories of codes as the level of analysis. For example, to explore how students made sense of the item, I first analyzed the frequency of codes across four categories: Overall Comprehension, Strategies for Comprehending, Read Aloud Deviations, and Corrections to Read Aloud. I conducted a series of two-way repeated measures analyses of variance (ANOVA) to investigate the statistical significance in the frequency of codes across the factors Linguistic Status, Presence of Illustration, and the interaction between Linguistic Status and Presence of Illustration. Additionally, I created radial graphs that visually compared the proportion of codes in each cluster across linguistic groups and presence of illustration. In these graphs a symmetric pattern across four quadrants indicates that the two groups compared have similar relative frequencies of codes observed across the illustrated and non-illustrated versions of items.

Second, at a micro level, I analyzed codes at the individual code level. For example, the number of cases for rereading an item, repeating a science term when reading the item aloud, or identifying a science term as not understood. Based on this level of analysis, I created symmetry graphs to examine pattern differences and similarities between. In these graphs, a pyramid shape indicates that the two groups compared have the same relative frequencies of codes observed.
In both the radial and symmetry graphs, codes are considered present or absent and do not consider multiple occurrences of the same code. For example, if a student identified that they did not understand two science terms in a given item, this was coded as present (value of a 1) to indicate a case of not understanding a science term for that item.

In a separate analysis, I accounted for multiple occurrences of the same code in the ways students made sense of items. If a student identified two science terms they did not understand, this was coded as a value of a 2. Next, I added all the actions coded for each student to calculate a total number of actions devoted to understanding the item. By comparing this total across student, I examined the magnitude of differences in the actions students from different linguistic groups carried out to make sense of illustrated and non-illustrated items.

Many researchers have speculated that ELLs devote more cognitive resources to making sense of the item than native English speakers (see Duran, 2008; Rivera, Collum, Wilner, & Sia, 2006). Utilizing this coding system provided in-depth information that compared the specific interactions of emerging bilinguals and mainstream students carry out when solving items. Next, I conducted independent t-tests to examine the statistical significance of frequency differences in coding categories across ELLs and Non-ELLs. Next, I performed dependent samples t-test to investigate the statistical significance of the frequency of differences in the coding categories across illustrated and non-illustrated items.

To investigate the differences in student performance based on (1) linguistic status, (2) presence of illustration, and (3) correct answer I investigated the percentage of students answering correct based on Presence of Illustration and Linguistic Status. Finally, I conducted two way repeated measures ANOVA to investigate the statistical significance of differences in student performance based on Linguistic Status and the Presence of Illustration.
Chapter 4

Qualitative Findings

This chapter presents the qualitative aspects of the students’ interactions with multiple-choice items. Starting with the two common items that all students answered, I discuss all three research questions. First, for each item, I provide a brief description of the test question and show the correct option with underlined text. Next, I include a table that lists the most frequent codes that describe how students made sense of the item, the strategies they used to solve the item, and how students reported using the illustration. During this discussion, I focus on the most frequent codes that were observed by students solving the question to highlight similarities and differences across linguistic groups and presence of illustration.

After presenting the qualitative results for each item in the study, in a separate section, I provide a detailed discussion of the bilingual participants’ use of Spanish during the cognitive interviews. Although the coding system was originally designed to analyze how students drew from Spanish throughout their verbalizations, less than a third of bilinguals used Spanish during their interview. This final section discusses their language use in detail.

Item 1

The first item, Seedling, was given to all students in its illustrated form. It is shown in Figure 3 below with the correct answer choice underlined. It assessed Life Science, a topic taught in Grade 7 at this school. The question is rather long and has complex syntax with a prepositional clause (of the large leaves) and two adverbial clauses (found on seedlings, growing in a forest).
What is the primary function of the large leaves found on seedlings growing in a forest?

A. To provide shade for the root systems  
B. To get rid of excess water that is entering through the roots  
C. To allow for leaf damage by insects  
D. To gather as much light as possible for photosynthesis

**Figure 3. Item 1: Seedling**

As shown in Table 7 below, less than half of emerging bilinguals correctly answered this item. In contrast, most mainstream students answered correctly. The majority of bilinguals who answered incorrectly should have learned this content, as seventeen of the twenty-two emerging bilinguals who incorrectly answered were in Grades 7 and 8. In contrast, almost half, three of seven mainstream students, who answered incorrectly were in Grade 6.

**Table 7**

*Item 1: Frequencies of Most Common Codes by Linguistic Status and Coding Category*

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>ELLs</th>
<th>n&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Non-ELLs</th>
<th>n&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered Correctly Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does not Understand Science Term</td>
<td>14</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sound Out Science Term</td>
<td>25</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitute Non-Science Term, Change Meaning</td>
<td>24</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitute Non-Science Term, Change Meaning</td>
<td>23</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall Fact</td>
<td>9</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process of Elimination</td>
<td>8</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guess</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illustration Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support Answer Choice</td>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus on Aspect in Stem</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not Use</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>Total n = 36
Making sense of the item. There were major differences in the ways students from different linguistic groups made sense of this item. Emerging bilinguals clearly carried out more actions to understand it, especially at the word level. As shown in Table 7 there were many more emerging bilinguals who reported they did not understand a science term, as 69% of bilinguals (25 of 36 students) reported that they did not understand at least one science term. This is compared to only six mainstream students who identified a science term they did not understand. In fact, Does not Understand Science Term was the most frequent code to describe how emerging bilinguals made sense of this item.

There were three science terms that emerging bilinguals commonly reported they did not understand. These terms were seedling, primary function, and photosynthesis with 17, 16, and 13 cases respectively. These terms accounted for 85% (46 of 54) of the total instances of not knowing a science term for emerging bilinguals. In addition, twelve emerging bilinguals did not know two of these three terms, and four students reported not understanding all three terms. This contrasts with only three mainstream students who reported not knowing what seedling meant, and two who reported not understanding what primary function meant. No mainstream student reported not knowing photosynthesis. Additionally, no mainstream student reported not knowing more than one of these terms.

Linguistic status and grade level did not appear to be associated with students reporting not understanding one of these science terms. There was an equal number of students from each grade level, about eight students from each of Grades 6, 7, and 8 who identified not understanding a science term. English Proficiency classification was also not associated with not understanding these terms. Students from all classifications of English proficiency (NEP, LEP, and FEP) identified these terms as unknown. Those students who stated they did not understand
all three of these terms included a Grade 6 student classified as LEP, a Grade 7 student classified as LEP, a Grade 7 student classified as FEP, and a Grade 8 student classified as NEP.

Surprisingly, not understanding a science term was not necessarily associated with choosing the wrong answer. As shown in Table 8 below, five out of nine emerging bilinguals who did not understand one of these words answered correctly. Of the twelve emerging bilinguals who reported not understanding two of these terms, five students answered the item correctly. Surprisingly, three of four ELLs who reported not understanding all three terms chose the correct option.

Table 8

*Table 8: Science Terms not Understood by ELLs*

<table>
<thead>
<tr>
<th>Number of science terms not understood</th>
<th>Number of students answering correctly</th>
<th>Number of students answering incorrectly</th>
<th>Total ELLs not understanding science terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

It appeared that the majority of emerging bilinguals who did not understand at least one of these terms but answered correctly relied on other terms in the item they understood to inform their answer choice. For example, one Grade 7 emerging bilingual classified as LEP reported that he did not understand all three of these terms. Notwithstanding, he was able to focus on the part of option D that he did understand and knew to be true, *to gather as much light as possible.* When recalling his steps for how he solved this item, he appeared to consider all options, but chose D. He stated, “I think it’s D cuz these are like saying that like they [the leaves] need a lot of sun to grow and they also need water and it kind of needs to grow so insects won’t damage it.”

His verbalization indicated that he knew plants need both water and sun to grow. During follow-
up questioning, I asked him if any part of the question helped him solve the item. He answered, "yah, the light, the light as possible.” In this case, the student was able to focus on the part of the item he knew to be true, that plants need light and water to grow. He chose this option even though he reported not understanding the term, photosynthesis.

Seventeen emerging bilinguals reported not knowing what a seedling was. This was compared to only three mainstream students who reported not knowing this term. Some emerging bilinguals focused on the root of the word, seed, to decipher its meaning. One emerging bilingual stated, “I just know it’s seed.” Although a seedling is related to a seed, this may have led students to focus on the fact that seeds are underground, whereas a seedling is a small plant above ground.

The second most frequent term emerging bilinguals reported not knowing was primary function, as 16 stated they didn’t understand at least one of these two words. Two bilinguals reported not knowing what a function was and three bilinguals reported not knowing what primary meant. Two other emerging bilinguals continually read this term as primary fraction a mathematical term. Not understanding this term made it difficult to understand what the question was asking. As a Grade 8 student classified as LEP put it, the most difficult part of this question was, "the primary function cause, like I don’t really get those two words, like that confused the whole question.” However, as previously discussed, this was not necessarily associated with an incorrect answer as just under half of those students reporting not understanding this term chose the correct option.

The final term, photosynthesis was identified by thirteen emerging bilinguals as unknown. An almost equal number, fourteen, sounded out this word when reading the item aloud. This contrasts with zero mainstream students who identified not knowing this science term, and only
two who sounded it out.

One notable strategy that emerging bilinguals used to comprehend other terms in this item was to draw from cognates. Although this was fairly infrequent, a few students were able to identify *excess* as a cognate. One Grade 7 student classified as FEP responded, “si porque es fácil porque es como exceso. (yes, because it’s easy because it’s like exceso)” Other terms that students identified not knowing had no cognate. For example, one Grade 6 bilingual classified as LEP, did not know what the non-science term *rid* meant in option B and he remained focused on the information in that option that he knew was needed for plant growth. He stated:

S: Well I know that it’s not to provide shade ‘cause that’d be sort of weird and it’s not to provide as much light as possible ‘cause maybe one day it could raining or something. And I think that’s what plants do, they get water to energize themselves, or something.

This student correctly explained that plants need water to grow, or energize themselves, as he stated. However, the option he chose stated *to get rid of water*, not obtain more water for the plant to grow. By focusing on the information he understood, he chose the incorrect option. In addition, this student reported not knowing what a *seedling* was. When asked what this item was asking, he described it by stating, “the function of the leaves found, like they’re growing on the forest floor.” While it first appeared that he interpreted *seedling* to mean leaf, he described that the picture helped him figure out it was a small tree. He stated, “I thought it [the seedling] was a leaf since it said leafs but then I saw it was this tree growing from the floor.” It appears that the illustration helped this student better understand the word he did not know, but the student remained focused on a plant’s need for water, choosing the distractor option B, which actually stated *to get rid of water*. 
One commonality across emerging bilinguals and mainstream students when they made sense of this item was that they frequently substituted non-science terms, changing the meaning of the substituted cases. For emerging bilinguals, these substitutions typically included prepositions (19 of 44 cases) whereas for mainstream students they typically included articles (18 of 28 cases), however prepositions were the second most frequently substituted words for mainstream students (9 of 18). It appears that, with respect to substituting non-science terms, there are some commonalities among emerging bilinguals and mainstream students.

Another frequent code to describe how mainstream students made sense of the item was to substitute science terms while maintaining the meaning. Typically, these included substituting the singular form of the word for the plural (root system for root systems), and did not affect the overall meaning of the item.

**Problem solving strategies.** Recall Fact and Process of Elimination were the two most frequently codes of problem solving strategies for both emerging bilinguals and mainstream students. However, there are more instances of each code for mainstream students than for bilinguals indicating that more mainstream students utilized these strategies. The most common facts that both emerging bilinguals and mainstream students recalled related to the necessary conditions for plants to grow, such as sunlight and water.

The third most popular strategy for emerging bilinguals was to make a guess. This code described when students explicitly stated that they guessed the answer. The third most frequent code for mainstream students was to choose the answer that “made sense.” This code described any time students reported that they chose an answer because it was logical, seemed right, or that they “just knew it.” In both cases, no specific strategy was observed or reported. Students who reported they chose an answer because it made sense could have in fact been guessing. However,
it could also be that these students knew the answer but could not articulate how they knew it or where they learned it.

**Use of vignette illustration.** Students from different linguistic groups were similar in the ways they reported using the illustration. Seven emerging bilinguals and five mainstream students reported not using the illustration, while nine emerging bilinguals and seven mainstream students were coded as using the illustration to support their answer choice. This code described cases of students justifying their answer choice by using an aspect of the illustration. Many times students interpreted the white spot in the illustration around the seedling as light. The white spot was originally included in the illustration to draw attention to the seedling rather than the larger trees in the background. However, throughout the course of the interviews, it became clear that this would be an aspect of the illustration that should be modified as it appeared to lead some students to pick the correct answer. For example, a Grade 6 emerging bilingual student provided the following verbalization during her think aloud:

S: Mm because D because the forest had like some spots of like sun and some plants grow there faster cuz it rains a lot there and they need some sun so it would be better if there were one spot of sunlight.

During her think aloud, she referred to both light and water as necessary for plant growth. It was not clear that the picture led her to choose option D, but when asked how she used the picture, it appeared that the illustration did influence her answer choice:

R: Did you use it [the picture] at all when you were solving the problem?
S: yah
R: Do you remember how you used it?
S: Yah, um I just, everything is shaded so like, except for this part so that’s where I got that the sun was in some spaces.

R: Ok, so that’s what got you to think there is sun in some places?

S: Yah

While it is possible that this student already knew this fact before seeing the picture, it is also possible that the illustration led her to pick the correct option.

Mainstream students reported using the illustration in a similar manner. When a Grade 7 monolingual student thought aloud, he did not mention the illustration:

S: Well, for a fact, I know leaves are for photosynthesis, and it makes sense that they’d be large when they were younger, because they’re small and they can’t get photosynthesis as fast as the larger trees around it, so I’d say it would have to be D, to gather as much light as possible for photosynthesis.

However, during the follow-up questioning, I asked him about the illustration:

R: Did you see the picture?

S: Oh, yeah, the picture helped, a little bit, yeah.

R: Do you remember how?

S: Well, I see the little sapling here, and it has, like, really big leaves and there’s, like, a little bit of light on it.

It appears that this student knew the answer and only used the illustration to help explain why he chose option D. However, it is possible that the illustration led him to this choice.

A major difference in students’ use of the illustration was that emerging bilinguals focused on an aspect of the item much more frequently than mainstream students. Many emerging
bilinguals used the illustration to focus on the seedling, or an aspect of the seedling. A Grade 7 student classified as LEP answered during the follow up questioning:

R: ¿Y piensas que usaste la imagen para entender la pregunta? [Do you think you used the image to understand the question?]

S: Pues sí lo usé un poco pero no mucho. [Yes, I used it a little, but not a lot.]

R: Pero no mucho, ¿y en qué pensaste cuando viste la imagen? [But not a lot, what did you think about when you saw it?]

S: Pensé que esta tiene una hoja más larga que los demás [I thought that it has one leaf that is longer than the rest].

Here, the student replied that she was able to see that one of the leaves on the seedling was larger than the other leaves. Unfortunately, it was not clear if she understood that this small plant was the seedling the item referred to. She also reported not knowing what primary function or seedling meant. In fact, she reported that a seedling meant, “muchos seeds” or many seeds. She incorrectly answered the question.

Surprisingly, the most frequently reported use of the illustration for mainstream students was that it helped them better understand the item. A Grade 6 monolingual stated that it clarified which leaves the item referred to. She explained the most helpful part of the item was:

S: Um, I liked that there is a picture there, because at first I didn’t know they were talking about, I thought they were saying large leaves on the bigger trees to help the little seedlings. But then I saw the picture and there was a little tree with huge leaves, so then I went back and read it again, so the picture helped.

Here, she was able to clarify that the item asked about the leaves of a small seedling rather than the leaves of larger adult trees. Nine mainstream students reported that it helped them better
understand what the question was asking while only two emerging bilinguals reported the same. One emerging bilingual reported that she used the illustration similar to the monolinguals’ previous use. The Grade 8 student classified as FEP stated:

S: Yeah, I was trying to figure out which one was the seedling, since I only saw the trees, the leaves up here, but then I saw that on the ground.

R: Do you think it helped you when you were solving it?

S: Yeah, ‘cause, like, it did because I knew which one—like, where it was and what it was and what it looked like, kind of.

She reported that the picture clarified what a seedling was by showing her where it was located and what it looked like. Unfortunately, this student answered this question wrong by guessing option B after she reported not knowing what photosynthesis meant. Another emerging bilingual, a Grade 8 student classified as LEP also reported the illustration helped him understand the item. He simply stated that the illustration helped because “it showed what they’re [the item is] meaning.” Although he chose the correction option, he also reported that the most difficult aspect of the item was “trying to find out what one [option] was best described by the picture,” indicating that the illustration may have influenced his answer choice.

**Combination tree.** Figure 4 provides the tree that traces the combination of actions for monolinguals responding to Item 1, including how students made sense of the item, the strategies they used to solve the item, and their use of the vignette illustration (VI). The goal of creating combination trees was to provide a clear visual representation of students’ steps to solving an item. As Figures 4 and 5 below show, students’ experiences solving this item greatly varied, which speaks to the uniqueness of each student’s set of approaches used to respond to the item. I created combination trees for each item in this study, one for bilinguals and monolinguals.
Figure 4. Combination tree: Bilinguals
Figure 5. Combination tree: Monolingual students
However, subsequent trees are not included as no overall trend can be identified using these representations. Similar complexities were observed for ELL students and for all items.

**Summary of findings for Item 1.** It was clear that emerging bilinguals carried out more actions to understand this item. In addition, they more frequently reported not knowing science vocabulary as compared to mainstream students. The most difficult terms for emerging bilinguals were: *primary function, seedling, and photosynthesis*. However, not understanding these terms was not associated with grade level or English proficiency classification. In addition, it did not always result in bilinguals choosing the wrong option. In many cases, emerging bilinguals reported not understanding one or more of these terms, but chose the correct answer by focusing on information they did understand. Unfortunately, focusing on information they understood was not always a successful strategy. Even though many bilinguals reported knowing that plants needed water and sunlight to grow, they often chose the wrong answer.

Monolingual and bilingual students were similar in the most frequent strategies they used to solve this item. A difference was that there were more cases of mainstream students using one or more strategies. It appears that emerging bilinguals carried out more actions to understand this item, whereas mainstream students carried out more actions to solve the item.

Finally, emerging bilinguals and mainstream students were both similar and different in their reported uses of the vignette illustration. One notable difference was using the illustration to better understand the item. Surprisingly, several mainstream students reported the illustration helped them clarify an aspect of the item whereas this was only the case for two emerging bilinguals. A commonality in the use of the vignette illustration was that students from both linguistic groups reported not using the illustration. However, many emerging bilinguals reported that they focused on an aspect mentioned in the stem when looking at the illustration. Finally,
this item calls attention to the need for careful construction of vignette illustrations as it appeared that the features of this illustration led some students to choose the correct option.

**Item 2**

All students in the study answered Item 2 with no illustration. It is shown in Figure 6 below. The item, nicknamed the Moon, assessed Earth and Space science, a topic covered in Grade 8. The linguistic features of this item are very different from those of Item 1. The overall length of both the stem and options are considerably shorter and the vocabulary consists of more frequent terms.

| The Moon produces no light, and yet it shines at night. Why is this? |
|--------------------|-------------------|
| A. The Moon reflects the light from the Sun. |
| B. The Moon rotates at very high speed. |
| C. The Moon is covered with a thin layer of ice. |
| D. The Moon has many craters. |

*Figure 6. Item 2: Moon*

Although the standard for this item was taught in Grade 8, it appeared to be fairly easy for students of all grade levels. As Table 9 shows, 67% of emerging bilinguals and 100% of mainstream students answered it correctly. Grade level did not necessarily predict which students answered this question incorrectly. All but two Grade 6 emerging bilinguals answered correctly, but almost half of Grade 7 emerging bilinguals (nine out of twenty) answered wrong. Only one Grade 8 emerging bilingual answered wrong.
### Table 9

**Item 2: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Category</th>
<th>ELLs</th>
<th>Mainstream</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered Correctly Making Sense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Substitute Non-science Term</td>
<td>24</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>- Sound Out Science Term</td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>- Substitute Science Term, Change Meaning</td>
<td>12</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- No Strategy</td>
<td>7</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>- General Background</td>
<td>7</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>- Knowledge</td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>- Apply Fact</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Process of Elimination</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Total n = 36

**Making sense of the item.** This item appeared to be fairly easy for both groups of students to understand. Similar to Item 1, substituting non-science terms was a frequent code to describe how students from both linguistic groups made sense of this item. Again, the vast majority of these substitutions entailed substituting articles and prepositions. One stark difference between mainstream students and emerging bilinguals is that 16 mainstream students inserted the article *a* when reading option B, “The moon rotates at *a* very high speed.” Only four emerging bilinguals carried out this specific substitution. Instead, emerging bilinguals greatly varied in the words they substituted, and no overall trend was observed.

Emerging bilinguals also frequently substituted science terms, again greatly varying in which science terms they substituted. The most commonly substituted word was *produces* (substituted by *products* or *process*) followed by *rotates* (substituted by *relates, rates,* and *reacts*). Emerging bilinguals were also coded as more frequently sounding out science terms. Eleven students sounded out the term *craters,* while other terms included *rotates* and *produces.* Finally, a
frequently used code to describe how mainstream students read the item aloud was omitting non-science terms (typically articles and prepositions). This was not a common occurrence for bilingual students.

None of the actions that students carried out to make sense of the item was associated with a correct or incorrect answer choice. In contrast to Item 1, very few students from both linguistic backgrounds reported not understanding terms or the entire item. Four emerging bilinguals reported not knowing what *crater* meant, and three of these students answered the item incorrectly. All mainstream students reported that they knew all the terms in this item.

**Problem solving strategies.** Students from both linguistic backgrounds used process of elimination and applied factual knowledge to solve this item. However, similar to the results discussed in Item 1, many more mainstream students were coded as carrying out these strategies than emerging bilinguals.

A successful strategy for many students from both linguistic groups was to apply factual information by describing the relative positions of the Sun, the Earth, and the Moon. A Grade 6 bilingual classified as LEP explained,

*S:* The moon produces no light and yet it shines at night. What, Why is this? The moon reflects the light from the sun. The moon rotates at every high speed. The moon is covered with a thin layer of ice. The moon has many cra-ters. I think it is A because the moon reflects the light from the sun because the sun is in the other side of the, um earth, and when it’s spinning, this, um rayos del sol [sunrays], the moon catches those rayos [rays]. And then it reflects. 

Here, the student explained that option A was correct because, as the Earth rotates, the Moon will reflect the rays of light that the Sun gives off. Although his explanation referred to one
specific location of the Sun, the Moon, and the Earth rather, he stated that the light will travel from the Sun, to the Moon, and then it reflects to the Earth.

Another successful strategy used by many mainstream students included eliminating distractor options based on their truthfulness or their relevance to producing light. Here, a Grade 8 mainstream student described how he eliminated the answers:

S: It’s because—I think it’s because the moon reflects the light from the sun, ‘cause I think that that’s, like—‘cause I know the moon doesn’t go very fast, because, like, if Earth doesn’t go very fast, then I don’t think the moon would either. And then, I don’t really think that ice is gonna help it shine much. Maybe a little, but it’s not—I don’t really think that’s gonna help. And then the moon has many craters, well, I don’t really understand how craters involves light.

The student eliminated answers by determining that some distractors would not produce light even if they were true. Many mainstream students used the process of elimination in combination with their background knowledge to solve the item. Here, a mainstream student drew from background knowledge to inform how he eliminated answers. He stated:

S: “The moon produces no light, and yet it shines at night. Why is this?” A. Well, first of all, I already kind of know this because of just, like, it’s background knowledge, but I’m still gonna ago through them. “The moon reflects the light from the sun.” B. “The moon rotates at a very high speed.” C. “The moon is covered with a thin layer of ice.” And D. “The moon has many craters.” And B is not right, because the earth does not shine, and I’m pretty sure it goes—it might go faster than the moon. C. “The moon is covered with a thin layer of ice,” where would the light still come from? Probably the sun. And, like, I
know on news that they found it underneath, but not, on top of the ground. And D, “The moon has many craters,” I know that’s true, but that wouldn’t make sense, so it’s A.

Combining process of elimination with background knowledge was also a successful strategy for emerging bilinguals. Eight of eleven ELLs who utilized process of elimination answered the question correctly. One FEP student in Grade 6 provided very concise reasoning for why she ruled out options B, C, and D. She stated:

S: Yeah. So B, it doesn’t really matter how fast the moon goes, ‘cause speed doesn’t produce light. And ice doesn’t produce light, and craters don’t really involve light at all.

However, this strategy was not successful for all students. The emerging bilingual below recalled her steps after choosing option D, the Moon has many craters:

R: OK. Could you tell me the steps that you took? ¿Los pasos que seguiste?

S: I read it, I read the options, and then I deleted some that I knew wasn’t possible.

R: Which ones were those?

S: B and C, ‘cause I kind of studied the topic, so I know the moon doesn’t have ice, and I didn’t want that, and these two, it was down to these two, and I don’t know what that means.

R: Craters?

S: And that one, “the moon reflects light from the sun,” I don’t get how it could do that.”

Here the student actually picked the answer that she reported not understanding. Process of elimination helped her narrow down her choices, but did not lead her to choose the correct answer.

Bilinguals who incorrectly answered this question varied in the options they chose. Similar to results in the first item, a few emerging bilinguals who answered incorrectly focused on true
information provided in an incorrect answer choice. For example, a student who chose option C, stated, “Um, I think it’s C, because I have heard that it has layers of ice.” While it is true that ice was recently discovered on the Moon, it does not have layers of ice. Another emerging bilingual focused on the fact that ice is shiny and related that to producing light, while yet another student reported that both the Moon and the Earth rotate, and decided on option B.

A difference between the strategies emerging bilinguals and mainstream students used was the number of students who referred to a classroom experience. Eight mainstream students utilized this strategy compared to five emerging bilinguals. This was surprising because, at the time of the interviews, the Grade 8 science teacher had just finished teaching about the moon cycles. All but one mainstream monolingual Grade 8 student reported that they learned about the content of this item in class. Not even half of Grade 8 emerging bilinguals, four of nine, recalled learning about it in class. When utilized, this was an extremely successful strategy. Every mainstream and emerging bilingual student who recalled learning this topic in class answered the question correctly.

Summary of findings for Item 2. Overall, students from both linguistic groups carried out fewer actions to make sense of this item as compared to the first item. Only a few emerging bilinguals reported not understanding the science terms crater and produced. These were also the terms emerging bilinguals most frequently sounded out.

While emerging bilinguals and mainstream students were similar in the type of strategies they used to solve this item, there were many more mainstream students using these strategies than emerging bilinguals. An interesting contrast across linguistic groups was that all mainstream Grade 7 students answered this item correctly whereas only about half of Grade 7 emerging bilinguals (ten of nineteen) answered correctly. It is not clear why this may be as those students
who answered incorrectly appeared to have made their choice for various reasons. However, a successful strategy for both mainstream and emerging bilinguals was to recall a classroom experience as all students answered correctly when utilizing this strategy.

**Item 3**

The first randomly assigned item, shown in Figure 7 below, assessed Scientific Processes, a topic covered throughout grades 6, 7, and 8. The text of the item is longer than most of the other items included in this study and contains unfamiliar vocabulary. First, it starts by referring to a Scottish scientist, Alexander Fleming, a Noble Prize winner who discovered penicillin. Second, the item includes the science term *agar*, a growth medium added to Petri dishes. Finally, this item is a cloze question which required students to finish the last sentence of the stem rather than answer a question.

Alexander Fleming noticed that bacteria growing on a plate of agar did not grow next to a mold that was growing on the same plate. He wrote in his laboratory report: "The mold may be producing a substance that kills bacteria." This statement is best described as

- A. an observation
- B. a hypothesis
- C. a generalization
- D. a conclusion

*Figure 7. Item 3: Alexander Fleming*

Overall, this item appeared to be very difficult for all students. As Table 10 below shows, only one emerging bilingual answered the item correctly in the illustrated form and one emerging bilingual answered correctly in the non-illustrated form. Mainstream students’ performance greatly differed based on presence of illustration. Three of four monolinguals answered the illustrated version correctly and no mainstream student correctly answered the version without an
### Table 10

**Item 3: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>Illustrated</th>
<th>Not Illustrated</th>
<th>ELLs</th>
<th>Mainstream</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answered Correctly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Sound Out Science Term</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>– Mispronounce Science Term</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>– Substitute Science Term, Maintain Meaning</td>
<td>5</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>– Does not Understand Science Term</td>
<td>5</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><strong>Problem Solving Strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Recall Fact</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>– Recall Classroom Experience</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>– Made Sense</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>– Process of Elimination</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>– Incorrectly Apply Fact</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>– Guess</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Illustration Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Focus on Aspect in Stem</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>– Visualize Aspect in Stem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

1. Total n = 3
2. Total n = 5
3. Total n = 4
4. Total n = 5
illustration. The correct answer to this question is option B, hypothesis. However, every student who answered incorrectly, regardless of linguistic background, chose option A. This causes great concern and indicates there might have been a systematic problem with the item. Ideally, an equal number of students should select each of the distractors.

**Making sense of the item.** One similarity across all four comparison groups was that students mispronounced the science term *agar*. Nine of the thirteen total students who responded to this item mispronounced this word. All students, regardless of linguistic status, appeared to be unfamiliar with this word. Surprisingly, only two emerging bilinguals and one mainstream student reported not understanding this term.

Students also had difficulties understanding this item at the syntactical level. Two emerging bilinguals reported that they did not understand the entire question and four mainstream students reported that the item was confusing. The illustration did not appear to facilitate understanding, as students taking the illustrated version also stated they did not understand it or found it confusing. The first sentence, perhaps intended to provide contextual information, appeared to hinder rather than help students. In order for students to correctly answer this question, they needed to be able to focus only on the second statement, and classify it as a hypothesis. This was not easy for students. As one Grade 7 mainstream student taking the item with an illustration put it, “I couldn’t really get, like, what I needed to classify as an observation, a hypothesis, a generalization, or a conclusion.”

However, although there were similarities, there were also many differences in the ways they made sense in this item. On the illustrated items, emerging bilinguals were frequently coded as sounding out science terms whereas on the non-illustrated item they more frequently reported not understanding science terms. These terms included *bacteria, agar, substance*, and all of the
options. In contrast, mainstream students frequently reread the item in both the illustrated and non-illustrated versions. Monolinguals also frequently repeated non-science terms when reading the item aloud.

**Problem solving strategies.** The most popular strategy across all comparison groups was to focus on an aspect of the stem. Nine of all eighteen students responding to this item used this strategy. However, four students (two ELLs and two mainstream students), focused on the first sentence, which states that *Alexander Fleming noticed bacteria*, a phrase that students associated with an observation. A mainstream Grade 6 student taking the non-illustrated item talked about the nuances in language that test takers must be aware of to correctly answer this item. During his think aloud he said,

S: Alexander Fleming noticed noticed that the bacteria growing on the plate, a plate of argar [sic] did not grow close to a mold that was growing on the same plate. He wrote in his laboratory report, ‘The mold may be producing a substance that kills bacteria.’ This statement is best described as: So this is a bacteria not growing next to mold, so the mold—maybe the mold—OK, so the statement is best described as an observation or hypothesis—OK. It’s not—he is observing that, but it’s not—he’s not, um—he doesn’t know that, so—and he’s saying “may,” so it’s more likely a hypothesis, generalization. He’s not really saying all mold, he’s saying the mold, so it’s not generalized. And a conclusion, maybe, but producing—it’s not solid, so I’m gonna say a hypothesis.

This student was able to focus on the essential information and pay close attention to the details of the language in this item. He described the need to be attentive to specific words, such as *may*. He also noted that the item did not say *all* mold, which could classify the statement as a generalization. When recalling his steps, he reiterates the need to pay attention to detail.
R: Can you recall your steps?

S: First I read it, then I picked out the important parts from it because it was kind of long and a little confusing. Not really, I just didn’t get it the first time, so I read it again. Then I looked at the possible answers, but I saw that A and B were kind of close, so I ruled out A. Then I looked at the rest of them and then I ruled out the other ones, so I knew it was B.

R: Did any part of the question help you solve it?

S: Um, [pause] maybe his—maybe, um, the mold may be producing instead of—and also the mold instead of mold is—the specifics there helped me narrow down what the possibilities were.

R: The mold helped you narrow it down?

S: Yeah. And also may be.

This student picked out what he called the “important parts” from the long item. These nuances were difficult to understand. Another mainstream student stated that the sequence of the sentences made it difficult to figure out which sentence the question asked about, he offered that,

S: Maybe the statement could have been at the front, like, what is his statement, so you know what you’re doing from the start, because I had to reread it after I knew that, because I hadn’t been listening for that in the question, so I had to reread it again.

Emerging bilinguals were similar to mainstream students in the strategies they used to solve this item. Many of them also focused on the first sentence when trying to solve the item. A Grade 7 emerging bilingual taking this item without an illustration originally chose the correct answer, hypothesis, but changed his answer to observation:
S: I think it’s B [hypothesis] because (silent for a few seconds). I think I changed my answer to A because he’s noticing the bacteria is growing on a plate of agar and it’s not growing next to mold.

Originally, the student correctly chose hypothesis, but upon second consideration he focused on the fact that Alexander Fleming was noticing, so he changed his answer to observation. During the follow up questioning, I asked him what the hardest part of answering this question was:

R: What was the hardest part of this question?

S: He wrote on his laboratory report

R: Ok, why was that hard?

S: Cuz it could be A and B

The student is clearly torn between choosing option A or option B. He expressed this by classifying the item as difficult because he believes either option could be the answer. This student is correct. A common practice in science is to write down observations of things one notices during an experiment. The last cloze sentence of the item ambiguously states, This statement is best described as. Students had to interpret this sentence as the second sentence, rather than the first. This distinction was difficult for students, as the majority believed the item was describing an observation.

Use of vignette illustration. Unfortunately, the illustration did not appear to help students better understand the question. Four students (one emerging bilingual and three mainstream) of the seven students taking the illustrated item, reported not using the illustration. One mainstream student and one bilingual student reported that the illustration helped them focus on a specific part of the item, such as the bacteria, or that Alexander was holding the plate. Another ELL
student stated that it helped him visualize the placement of the mold in relation to the bacteria. However, in each of these cases, the students chose the incorrect option.

**Summary of findings for Item 3.** All students had difficulty understanding what this item was asking them to do. More emerging bilinguals identified science terms they did not know whereas more mainstream students reread the item to understand it. Both emerging bilinguals and mainstream students focused on aspects of the item to solve it, but this strategy was associated with choosing a wrong answer. The illustration did not appear to help students understand the item. Most mainstream students reported not using the illustration, two of three emerging bilinguals reported that they used it to focus on or visualize an aspect of this item, but chose the wrong answer.

**Item 4**

The second randomly assigned item 4, shown in Figure 8 below, was classified as assessing Scientific Processes, a topic covered in Grades 6, 7, and 8. The item asks students to correctly order the steps to using a stopwatch.

Maria wanted to measure the amount of time it took for a ball to roll down a ramp. She had never used a stopwatch before. Kevin gave her the following directions, but they were in the wrong order.

Step 1: Hold the stopwatch in one hand.  
Step 2: Press the button once to start the clock.  
Step 3: Press the button twice to clear any old times.  
Step 4: Press the button to stop the clock.  
Step 5: Let the watch run until it is time to stop it.  
Step 6: Record the amount of time.

How should she arrange Kevin’s steps so they are in the correct order?
While the text of the item was long, only two words were identified as science terms. Student performance was fairly consistent across linguistic groups and presence of illustration. About half of the emerging bilinguals answered the question correctly in both the illustrated and non-illustrated versions. About half of the mainstream students answered the question correctly with an illustration, and all of them correctly answered the version with no illustration.

Table 11 below summarizes the most frequent codes for making sense, problem solving strategies, and use of illustration. All students taking the item, regardless of illustration, substituted non-science terms. Emerging bilinguals substituted words that maintained the original meaning when the item was illustrated (e.g., your hand instead of one hand). Emerging bilinguals’ substitutions on the non-illustrated version changed the meaning of the text. To inform their substitutions, they appeared to be drawing from the visual information of the word. These substitutions included hold the stopwatch in on hand, and clear any cold times. In both cases the substituted word closely resembled the printed text. Mainstream students greatly varied in the types of substitutions they carried out, including, the for she, and stopwatch for stopwatch. It appears that most of monolinguals’ substitutions were also based on visual information.

One difference in the ways students made sense of this item was that emerging bilinguals repeated non-science terms, and monolinguals did not. There was no obvious trend in the words that emerging bilinguals repeated, including they, correct order, roll, let the watch run, and it.
Table 11

**Item 4: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>ELLs</th>
<th>Mainstream</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Illustrated</td>
<td>Not Illustrated</td>
</tr>
<tr>
<td><strong>Answered correctly</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Making Sense</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitute Non-science, Maintain Meaning</td>
<td>Illustrated</td>
<td>2</td>
</tr>
<tr>
<td>Reread</td>
<td>Illustrated</td>
<td>2</td>
</tr>
<tr>
<td>Repeat Non-science Term</td>
<td>Not Illustrated</td>
<td>2</td>
</tr>
<tr>
<td><strong>Problem Solving Strategies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Made Sense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background Knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Made Sense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Illustration Use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualize Aspect in Stem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not Use</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Total n = 5
2 Total n = 4
Two monolinguals solving this item described it as confusing. One mainstream student identified Step 3 as confusing, which reads, “Press the button twice to clear any old times.” He interpreted *old times* to mean past times, or in the olden days. Another mainstream student stated that having the steps numbered from 1-6, but then asking the right order was “difficult to process.”

Only one emerging bilingual identified the steps as confusing, stating, “Just the steps, the way it was put in these, I gotta think about it in a different way.” This statement indicates he may have described using a stopwatch in a different way. Other students could not specify an exact reason why the item was confusing. As a Grade 6 mainstream student taking the item with no illustration stated during the follow up questioning:

R: Was anything confusing?

S: “Maybe some of, like, the steps [are confusing], it was kind of a little, like, “What?”

R: (Like which one?)

S: I can’t find it now. Let the watch run till it’s time to—oh, no, that’s not it. Press the button twice to clear any old times. When you’re reading through the list, that’s like, “Huh?” But if you reread it slowly, it’s like, “Oh!”

Overall, this item did not appear to be too difficult for students to understand. Students understood the vocabulary and knew what the question was asking. Only a few students found it confusing.

**Problem solving strategies.** It was difficult to observe explicit strategies that students used to solve this item. Many students just repeated back the order of steps they thought was correct as their strategy. Four students reported that they chose the sequence that “made sense” or “seemed right,” and this strategy was typically associated with the correct answer choice.
Another successful strategy observed for both one emerging bilingual and two mainstream students was to project one’s self onto the item. These students stated that they chose the answer reflecting how they would use a stopwatch. A Grade 8 emerging bilingual taking the item with no illustration stated,

S: So I think it’s, one, two, five, four, uh I think it’s D. Cuz, like I think, cuz that’s the way I would do it. I would do it in those steps, and then um yah I just woulda said it in that order.

The student stated that he chose an answer that matched how he used stopwatches. This type of verbalization was similar to the two mainstream students taking the item with no illustration who also reported that they chose their answer because it was “how I would use it.”

**Use of vignette illustration.** Students from different linguistic backgrounds greatly differed in the ways they used the illustration on this item. While the three bilingual students taking the illustrated item were not coded as using any observable strategy, they verbalized that the illustration helped them visualize aspects in the item. One student stated, “I imagined I was holding a stopwatch.” While another said that he “thought of a person actually doing it” when he saw the illustration. These responses align with other students who were coded as projecting themselves onto the item to find an answer. Although one mainstream student described the illustration as “cheesy,” he went on to state that the illustration helped him. He stated, “It kind of helped me to pick the first one. I knew the 1’s were right, because it was, like, holding, the person was holding it in their hand.”

However, the majority of mainstream students, three of four, did not use the illustration and stated that it was not helpful. As one monolingual put it, “It just shows what—if someone didn’t know what a stopwatch was, they could look at that and figure out what it was, I guess.” This was compared to only one emerging bilingual who reported not using the illustration and stated,
“well, it [the item] just kind of told me what the stopwatch did and how to use it, so I didn’t really need the picture.”

**Summary of findings for Item 4.** This item did not appear to be difficult for students to understand. The most frequently observed actions to make sense of the item included substituting non-science terms for both emerging bilinguals and monolinguals. However, emerging bilinguals were more frequently coded as repeating non-science terms whereas mainstream students reread the entire item. More often than not, no specific strategy was observed to describe how students solved this item. The most common strategy observed was self-projection. Finally, there were significant differences in the ways bilinguals and monolinguals reported using the illustration. Three of four mainstream students reported not using the illustration. In contrast, three of four emerging bilinguals reported using it to visualize an aspect related to solving the item.

**Item 5**

Shown in Figure 9 below, Item 5 assessed the standard, Physical Science, and asked students to explain why helium balloons float, a content taught in Grade 6. This item was much more difficult for emerging bilinguals than it was for mainstream students. The majority of bilinguals incorrectly answered this item, whereas all mainstream students answered correctly.

A balloon filled with helium gas is set free and starts to move upward. Which of the following best explains why the helium balloon moves upward?

A. The density of helium is less than the density of air.
B. The air resistance lifts the balloon up.
C. There is no gravity acting on helium balloons.
D. The wind blows the balloon upward.

*Figure 9. Item 5: Helium Balloon*
**Making sense of the item.** This item did not appear difficult for students to comprehend, and very few students identified terms they did not understand. Table 12 below shows that emerging bilinguals substituted both science and non-science terms more frequently than mainstream students. When substituting science terms, they changed the meaning of the original word. For example, three emerging bilinguals substituted *destiny* for *density* on both the illustrated and non-illustrated versions. Another frequent substitution was for the word, *resistance*, which emerging bilinguals substituted as *restains, distance*, and *restates*.

There was not more than one observation for any given code to describe how mainstream students made sense of this item. That is, mainstream students read the item very closely to the printed text, did not utilize an observable strategy to understand it, and reported no difficulties understanding any terms or phrases. For this reason, there are no codes listed in Table 12. The only coded used more than once to describe how mainstream students made sense of the non-illustrated version of the item was inserting a non-science term. This was also a frequent code for emerging bilinguals taking illustrated and non-illustrated items. Students from both linguistic backgrounds inserted articles when reading aloud. A common insertion observed by both monolinguals and bilinguals was to insert *the* into option C, “There is no gravity action on *the* helium balloons.” Two emerging bilinguals also inserted *the* into option A stating, “The density of *the* helium is less than the density of air.

**Problem solving strategies.** There was great heterogeneity across and within linguistic groups for students’ problem solving strategies. I identified eight strategies that emerging bilinguals carried out to solve the illustrated and non-illustrated versions and eight different strategies that mainstream students used on the illustrated form. Unfortunately, the majority of strategies that bilinguals used were unsuccessful as only two bilinguals chose the correct option.
Table 12

**Item 5: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>ELLs</th>
<th>Illustrated</th>
<th>Not Illustrated</th>
<th>Mainstream</th>
<th>Illustrated</th>
<th>Not Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered Correctly Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Insert Non-Science Term</td>
<td>Repeat Non-Science Term</td>
<td>Insert Non-Science Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Substitute Non-science Term, Maintain Meaning</td>
<td>Repeat Non-Science Term</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sound Out Non-science Term</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Substitute Science Term, Change Meaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Substitute Science Term, Change Meaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Partial Reasoning</td>
<td>Incorrectly Recall Fact</td>
<td>Recall Fact</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Self Evaluate</td>
<td>Made Sense</td>
<td>Process of Elimination</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Self Project</td>
<td>No Strategy</td>
<td>Classroom Experience</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Process of Elimination</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Incorrectly Recall Fact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Self Monitor</td>
<td></td>
<td>Classroom Experience</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Focus on Aspect</td>
<td></td>
<td>Everyday Experience</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Eliminate Answer</td>
<td></td>
<td>Did not Use</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Illustration Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Made Sense</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

1 Total n = 5
2 Total n = 3
3 Total n = 4
Half of the emerging bilinguals who chose the wrong option chose C, *there is no gravity acting on helium balloons*. However, it was difficult to discern exactly why these students chose this option. One bilingual stated, “I look at the picture and I see the balloon going so I would say that C, that there is no gravity acting on helium balloons and the helium just takes the balloon up.” Similar to the other two emerging bilinguals who chose this option, this student focused on the balloon traveling upward without acknowledging the presence of gravity.

Two other emerging bilinguals were coded as using partial reasoning; one student answered incorrectly and one student answered correctly. The first student, a Grade 7 bilingual classified as LEP answered correctly stating, “Um [option] A because helium makes the balloons goes up and then, um, (pause) and it moves up upward (pause) it moves up with helium, with just helium, not with any other kind of air.” This answer is correct, that helium is a “special” kind of air, because it is less dense than oxygen. However, hydrogen is also less dense than oxygen and would cause balloons to float up. So the statement that, “not with any kind of air” was not completely accurate and was coded as a partially correct answer. In addition, this student did not refer to the density of helium in relation to air. It appears that this student was one the right track, but did not provide a complete rationale.

Two emerging bilinguals chose option B, *air resistance lifts the balloon upward*. It appeared that these students did not understand both the question and the relationship among the forces and objects described in the item. A Grade 6 student classified as LEP provided the following verbalization:

S: A balloon filled with helium gas is set free and starts to move upward. Which of the following best explains why the helium balloon moves upward? A, the density of the helium is less than the density of air, B the air resistance lifts the balloon up, C there is no gravity
acting to helium balloons, D the wind blows the balloon upward. Um, usually the wind
doesn’t blow from bottom, it blows from side to side, so it can’t be D. And I’m guessing
there is not that much, but a little bit of gravity in the balloon, so it’s not C, so it’s either A or
B. Um, [reading quietly] I’m guessing that helium is stronger than air, so is—it can’t be that
one, because this one’s less, and I think it’s more, so it’s gonna be B, since it’s the last one.”
R: Could you tell me the steps that you took?
S: Yeah, um, I eliminated the ones that can’t be, for sure it can’t, because wind can’t blow
from bottom, and there could be gravity there, so that probably won’t work. And then, like,
probably helium is, like, more like—is less than the density of air, so it only leave B.
This student acknowledged the correct option as true then he stated, “helium is, like, more like--
is less than the density of air.” However, he also stated that helium is stronger than air, which he
contradicted with less dense, and ruled out the correct answer choice.

In contrast to emerging bilinguals, this item appeared to be very easy for mainstream
students. Successful strategies for monolinguals included referring to the periodic table of
elements, experiments in class, or a related concept: water floating. One student noted that their
teacher may have used this exact example in class to teach density. He stated,

S: I think it’s A, because we learned about density, and that’s actually kind of one of the
examples, I think, that we had, and we’re also studying the periodic table of the elements
right now, and I know that helium is less dense than air, because it’s the second less dense
element of the periodic table.”

Here, the student focused on the fact that the item asked about the density of helium in relation to
air. The example used in the item was helpful to him because it replicated an activity used to
teach this concept in class. Not one emerging bilingual referred to a classroom experience when solving this item.

**Use of vignette illustration.** Students across linguistic groups greatly differed in the ways they reported using this illustration. Four of five emerging bilinguals used the illustration to focus on an aspect in the stem. Two of these students focused on the fact that the balloon was floating upward, a third student focused on the fact that the balloon was being let go. This bilingual also used the illustration to project himself onto the item. He stated, “once I had a balloon I thought that, I let it go and it went up cuz the wind was strong.” This student related the illustration to his own previous experience of the wind blowing and taking away his balloon. He did not discuss helium in his rationale and appeared to be focused on his own experience with balloons rather than the situation described by the item.

Similar to Item 4, Stopwatch, three of the four mainstream students reported not using this illustration. As one mainstream student stated, “the explanation kind of just said it all” indicating that the illustration wasn’t necessary. The only mainstream student who reported using the illustration, stated that, “It just kind of showed a balloon floating as a person let it go and not, like, dropping or any of that stuff. And yet again, it helped give me—it helped reboost my background knowledge.” The illustration may have served a motivational function for the student to help him recall previous knowledge.

**Summary of findings for Item 5.** Overall, students from both linguistic backgrounds did not carry out many actions to make sense of this item. However, as with previous items, emerging bilinguals were coded as carrying out more actions to make sense of this item compared to mainstream students. There was heterogeneity in the strategies students used to solve this item. Unfortunately, most of the strategies emerging bilinguals used were not successful, as the
majority of bilinguals incorrectly solved this item. In contrast, all mainstream students successfully solved this item. Finally, most emerging bilinguals reported using the illustration to focus on an aspect of the item, but this did not lead to correctly answering the question. In contrast to emerging bilinguals, the majority of mainstream students reported not using the illustration.

**Item 6**

Item 6, in Figure 10 below, also assessed Physical Science, a Grade 6 content. Most students, regardless of linguistic background, incorrectly answered this question. As shown in Table 13, no emerging bilingual correctly answered the illustrated version of the item, but three of five correctly answered the item without an illustration. Mainstream students’ performance was more consistent across items, as two of five monolinguals correctly answered in each case.

<table>
<thead>
<tr>
<th>A piece of pine wood floats on the surface of a lake because the water exerts</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. an upward force equal to the weight of the wood.</td>
</tr>
<tr>
<td>B. a downward force equal to the weight of the wood.</td>
</tr>
<tr>
<td>C. an upward force equal to the weight of the displacement water.</td>
</tr>
<tr>
<td>D. a downward force equal to the weight of the displacement water.</td>
</tr>
</tbody>
</table>

*Figure 10. Item 6: Pinewood*

**Making sense of the item.** The linguistic features of this item challenged all students. It was a cloze question with unfamiliar science vocabulary. Emerging bilinguals reported not knowing science terms more often than mainstream students. Eight of the ten emerging bilinguals reported that they did not know what *exert* meant, including the three students who correctly answered the item. In contrast, only two mainstream students reported not knowing what *exert* meant. Two
Table 13

**Item 6: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Category</th>
<th>ELLs Illustrated</th>
<th>ELLs Not Illustrated</th>
<th>Mainstream Illustrated</th>
<th>Mainstream Not Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered Correctly Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Does not Understand science Term</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>- Substitute Science Term, Change Meaning</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>- Substitute Science Term, Maintain Meaning</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>- Omit Non-science Term</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Classroom Experience</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>- Apply Fact</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>- No Strategy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>- Focus on Aspect in Stem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Focus on Entire Stem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Visualize Aspect in Stem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Connect to Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illustration Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Focus on Aspect in Stem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>- Focus on Entire Stem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Focus on Aspect not in Stem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Visualize Aspect in Stem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Connect to Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Total n = 5
emerging bilinguals and two mainstream students reported not knowing both of the terms *displacement water* and *downward*, and answered the item wrong.

When emerging bilinguals made sense of the illustrated item, they frequently substituted science terms, changing the meaning. Substitutions included substituting *downwater* for *downward*, *height for weight*, and *displant for displacement*. In contrast, monolinguals rarely substituted words on the illustrated version of the item. Emerging bilinguals also repeated and sounded out the science terms *exerts, displacement, upward, and downward* when reading the item aloud. These terms appeared to be unfamiliar for emerging bilinguals.

Mainstream students identified this question as being difficult because it was a cloze question that required them to finish a statement. A Grade 7 mainstream student, who took this item with an illustration, expressed her confusion as she thought aloud and recalled her steps. She stated,

S: A piece of pine wood floats on the surface of a lake because the water exerts, an upward force—Hmm. I’m not really sure what this is asking, like, it doesn’t really seem like a question. But a piece of pine wood floats on the surface of a lake because the water exerts…

Here, as she thought aloud, she stated that she didn’t see a question. When she recalled her steps, this confusion became explicit:

S: I had to read this a couple times, because I wasn’t sure if it was a question, but it is, because it’s, like, saying, because the water exerts—it’s—like, you have to answer the rest of the question. So it was a little confusing. I would, if I were to write this, I’d rewrite this to make it more of a question, and I’d put, like, dots, so that you could, like, tell that you have to answer the rest of the question, too. And then I just went through, I had to read these a couple times and I had to, like, figure out which was most, like, logical, and so I just chose D because it dealt with the weight and the force.
During follow up questioning she reported that the most difficult aspect of solving this item was that it, “wasn’t a question, but rather required you to fill in the sentence.” In all three parts of her interview, she reported being confused by the item’s format. Surprisingly, no emerging bilingual reported that this was a confusing or difficult aspect to the item. Rather, bilinguals were much more frequent to report that specific words were difficult because they didn’t know their meanings.

**Problem solving strategies.** Students greatly differed in the strategies they used to solve this item. Most students who incorrectly answered chose option A, *upward force equal to the weight of the wood*. Of the seven emerging bilinguals who answered incorrectly, five chose option A. Additionally, of the seven mainstream students answering incorrectly, six chose option A. The unfamiliar term *displacement water* appeared to have turned students off from choosing the correct option.

A frequent strategy for emerging bilinguals was to recall classroom experiences, with three of five emerging bilinguals invoking this strategy. Unfortunately, this strategy was associated with a wrong answer in all cases as the unfamiliar science terms made it difficult for emerging bilinguals to relate their previous knowledge to the correct answer. A Grade 8 student classified as LEP, reported not knowing *exert*, which distracted him during problem solving. He points out the importance of understanding each word in the item, and the result of not knowing this term meant he was not able to answer the item:

*S:* A piece of pine wood floating floats on the surface of the lake because the water ex ex extentret, extends, I don’t know what that word is, I just don’t know what this word means […] an upward force equal to the weight of the wood, the downwater force equal to the weight of the wood, the upward force equal to the weight of the dis-pla-ment water, a
downward force equal to the weight of the dis-ment water. (rereading to himself) Um, well I think it’s A because it’s telling me I think it’s telling me the water, the wood is less weight than the water so that the water can hold it up and the wood can’t really sink down.

R: Ok, and could you tell me the steps, los pasos que seguiste?

S: Well, I was thinking about, well, what I did in science, that um, like if it’s more dense it goes down and if it’s less dense it goes up so, um I couldn’t figure out that word …

R: Which one?

S: The…… the one in the question

R: In the question?

S: yah in the question so I was just saying if it is less dense then it goes upward.

R: Ok,

S: So the water of the force, it just goes up.

R: Did any part of the question help you solve it?

S: Well the only thing that helped me was like in sixth grade I remember a little bit like when it is water and when it’s water and wood that a lot of times, that like wood a lot of the time, like most of time, wood always floats, ‘cause it’s less dense than the water.

R: Ok, and ¿qué fue la parte más difícil? What was the hardest part?

S: Uh, the question

R: The question? Anything specific? ¿Algo especifico?

S: The water, the word.

R: Exerts?

S: Exerts, yah
In his verbalizations, the student explained that not understanding *exert* prevented him from understanding the overall meaning of the question. Even though he remembers learning in science that objects that are less dense will float, and objects that are denser will sink, the linguistic features of the item prevented him from applying what he learned in class and showing that he knew the concept being assessed.

In contrast, emerging bilinguals taking the non-illustrated version never recalled a classroom experience. Strategies of bilinguals correctly answering the illustrated version included: guessing, visualizing, and provided partially correct reasoning. However, it appeared that even those students who answered correctly still found this item difficult to understand. A Grade 8 student classified as LEP thought aloud:

*S: Uh, so I think it’s C, because most of the things float on the water because it’s equal to, like it’s like an equal force that they have, so yah.*

*R: And then can you just recall the steps?*

*S: I read the the what’s it called?*

*R: the question?*

*S: yah, the question, and then I didn’t get it and I read it again like three times then I read the choices and then I read the question again and then I started thinking about the, um, why things float and then I thought it was they have equal force, yah.*

During the think aloud, the student began to explain that most things float because of the equal forces. However, during his recall, he stated that he read the item three items to try to better understand it. Fortunately, he focused on the fact that he knew objects that have equal force will float, the force of the water will be upward, and chose C.
Similar to emerging bilinguals, mainstream students also reported a variety of problem solving strategies for this item. Across illustrated and non-illustrated items, mainstream students used process of elimination, guessed, recalled everyday experiences, focused on important information in the stem, and applied factual knowledge. When solving this item, many students reported that the term *displacement of water* was unfamiliar. A Grade 8 student provided partially correct rationale,

S: B and D are out because *up. Displacement of the water* offers a bit of confusion. Wait a minute. [pause] I think it’s A.

The student identified the term as confusing in his think aloud. He confirmed this confusion when he answered what the hardest part of the item was, “displacement of water, I guess either I forgot it or never learned it because of different school curriculums or something.” Other monolinguals reported similar confusion with this term. The result was most students picking option A, which had similar meaning, but does not contain this unfamiliar term.

**Use of vignette illustration.** There was a large variety of ways that all students reported using the illustration. These uses included focusing on various aspects of the stem, making connections to science topics, and visualizing. Some students used the illustration to find information not provided in the stem. A Grade 6 LEP student used the illustration to see exactly how the log floated. She stated,

S: I’d say about, um let’s see, A. Because um, like if it’s the same, if it weighs the same as the water, that would be like in the middle of the water. Downward it would be tying it down but the top is still showing [in the illustration] and A, it would be more showing than the downward.
Here, the illustration helped the student know exactly how the wood floated. Because she saw that part of the log was above water, she knew that the force must be upward. While she chose the incorrect answer, A, she showed that she understands the relationship between the forces. This is similar to, a Grade 6 mainstream student who also looked at the illustration for additional information:

S: A piece of pine wood floats on the surface of a lake because the water exerts: Upward force equal to the weight of the wood. A downward force equal to the weight of the wood. An upward force equal to the weight of the displacement of water. [pause] Hmm. OK, so you can see the ripples. I would say C.

Later, when I ask what helped her solve the problem she reported,

S: Definitely the picture.

R: OK. How did the picture help you?

S: The ripples in the picture. It kind of made me think that it would be the upward force equal to the displacement of the water, since the water is moving out of the way. So that really helped.

The student used the picture to focus on the fact that the water is displaced. In general, this illustration seemed to help students from both linguistic groups solve the item. Both emerging bilinguals and mainstream students reported using the illustration to focus, visualize, or make a connection.

**Summary of findings for Item 6.** Emerging bilinguals and mainstream students were similar in many aspects of solving this item. Both groups of students identified not understanding science terms, including exert and displacement of water. These unfamiliar terms often resulted in students selecting a distractor option that they better understood. Students from both linguistic
groups utilized a wide variety of strategies to answer the item, and there was no clear trend of which strategies were associated with the correct answer choice. Finally, both emerging bilinguals and mainstream students reported using the vignette illustration to focus, visualize or make connections to science.

**Item 7**

Item 7 assessed Physical Science, taught in Grade 6, and required students to understand the changing states of matter. More specifically, students had to know that when two elements chemically combined they created a compound. This was a difficult item for emerging bilinguals. As Table 14 shows, zero of three bilinguals answered the illustrated version correctly and two of four bilinguals correctly answered the item with no illustration. This item appeared to be slightly easier for mainstream students as three of four students correctly answered the item with an illustration and two of four correctly answered the item with no illustration.

When magnesium (Mg) metal is burned in the presence of oxygen (O2), magnesium oxide (MgO) is produced. The properties of magnesium oxide are different than the individual properties of magnesium and oxygen because magnesium oxide is

A. a solution.
B. a mixture.
C. a compound.
D. an element.

![Figure 11. Item 7: Magnesium oxide](image)

**Making sense of the item.** A common observation across emerging bilinguals and mainstream students making sense of this item was eliminating the initials that represented the elements and compounds when reading the item aloud. Seven mainstream students and two emerging bilinguals omitted these abbreviations. Emerging bilinguals frequently mispronounced...
Table 14

*Item 7: Frequencies of Most Common Codes by Linguistic Status and Coding Category*

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>ELLs Illustrated</th>
<th>ELLs Not Illustrated</th>
<th>Mainstream Illustrated</th>
<th>Mainstream Not Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered Correctly Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>− Does not Understand Science Term</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>− Sound Out Science Term</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>− Mispronounce Science Term</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>− Incorrectly Apply Fact</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>− Process of Elimination</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>− Guess</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>− No Strategy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Illustration Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>− Did not Use</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>− Visualize Aspect in Stem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>− Visualize Entire Stem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>− Focus on Aspect in Stem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

1 Total n = 3
2 Total n = 4
words including *magnesium, oxide, individual* and *presence*. In addition, many emerging bilinguals reported not understanding science vocabulary. Three emerging bilinguals taking the illustrated version of the item identified not knowing the meaning of *magnesium*, while two reported not knowing the element abbreviations (e.g., MgO). In all cases, bilinguals who reported not understanding these terms answered the item wrong. Although these terms have Spanish cognates, only one emerging bilingual referred to *magnesio* and *oxígeno*. In contrast, only one mainstream student identified not understanding *magnesium*.

**Problem solving strategies.** Similar to previous items, all students utilized a wide variety of strategies across linguistic groups and presence of illustration. One commonality across linguistic groups was that students recalled factual information. Typically these facts consisted of students defining the different answer choices (*mixture, element, and compound*). For example, a Grade 7 mainstream student reported during his think aloud:

S: I think a mixture, because it’s two—no, actually, no. It’s a—either an element or a compound. Since those are both elements, I’d probably say it was a compound, since I don’t know, it’s been a while since physical science, but I think two elements combined don’t make another element, so it’s probably a compound. A mixture would be when you usually combine, like, two powders or liquids together and they kind of mix together, and a solution is where one is a solvent and the other is a solute, which means that one gets solved, like salt and water. The salt—it becomes a salt water solution, the salt is dissolved.

This student defined all four answer options which allowed him to confidently eliminate distractors and select the correction option, *compound*. Similarly, an emerging bilingual also defined compound to correctly solve the item:
S: Well, we did this experiment with sulfur and iron, and since magnesium and oxygen are elements, they can’t make an actual element, so then it’s not a mixture because they’re chemically bonded, and they’re not a solution because it—um, um, dang it, I forgot. Well, um, they’re not a solution because a solution is, like, mixing two things together but it can still separate. A compound is when you mix—two things are mixed chemically bonded.

These examples illustrate how students from both linguistic groups successfully applied the definition of a compound to this item to correctly answer the question.

Three emerging bilinguals focused on the two elements described in the item as coming together and mixing, which they associated with option, mixture. A Grade 7 LEP student provided the following verbalization:

S: I think it is a mixture cuz I think it has two, like two elements and they mix, and I think it’s B.

R: Ok, great. And then can you tell me the steps that you took? Los pasos que seguiste?

S: I looked at all of…

R: at what?

S: I looked at mag-zea-zeum and I looked at the A,B, C and D and I thought it was a mixture cuz it has like two kinds of names.

In this example, the student focused on the fact that the item presented two different types of names, but does not go as far as noting they are chemically bonded.

Another emerging bilingual, A Grade 8th NEP student recalled learning about the purpose of compounds from his father, which helped him correctly answer the question:

S: Creo que es un componente, like mi papa me dijo la otra vez y creo que es un componente porque el magnesium se puede usar like para muchas cosas. [I think it’s a component, like
my dad told me the other time, and I think it’s a component because magnesium can be used for many things.]

Although this student does not provide a definition, he reported a discussion between his father and him about magnesium and its many uses. While he incorrectly translated the term from English to Spanish, *compound* as *componente* (component), he still chose the correction option.

**Use of vignette illustration.** The illustration did not appear to be helpful for most mainstream students and emerging bilinguals. The majority of students, one of three emerging bilinguals and two of four mainstream students, reported not using the illustration. In addition, a third mainstream student reported that the illustration was confusing. He described:

S: Uh, I don’t know. The picture didn’t make a lot of sense. It’s like a stick being held above—probably like a stick of magnesium being held above a flame, oxygen. Fire sucks in oxygen, so I don’t know why fire would matter, but I don’t know.

Here, the illustration conflicted with the student’s understanding of how the fire and magnesium interact to produce magnesium oxide which resulting in his statement that the illustration did not make sense.

Two students, one emerging bilingual and one mainstream student, reported using the illustration to focus on the entire item. A Grade 6 mainstream student described,

S: I glanced over to it a couple of times so that I would make sure that I knew that magnesium metal would be a solid, whereas there were lines on top of the metal, so it made it look like a gas. So when that gas is mixing with the oxygen, it would make magnesium oxide. So that kind of helped put the entire thing together.

Two other emerging bilinguals reported that the illustration helped them visualize an aspect of the item. A Grade 7 student classified as LEP explained that he used the illustration to see
aspects of the item, “vi que, como es metal y está burned, like in the presence of oxygen, yo creo. Yo creo que it produces MGO.” [I saw that it’s metal and it’s burned in the presence of oxygen, I think. I think it produces MGO.] Although the majority of mainstream students reported not using the illustration, it helped two of three bilinguals visualize or focus when solving the item.

**Summary of findings for Item 7.** Both differences and similarities existed in the ways that emerging bilinguals and mainstream students interpreted this item. Students from both linguistic groups sounded out and omitted science terms. However, emerging bilinguals identified more science terms they did not understand, and also mispronounced science terms more frequently than mainstream students. Students carried out a wide variety of strategies across linguistic group. One similarity was that emerging bilinguals and monolinguals recalled facts to successfully answer the item. While most mainstream students reported not using the illustration, two of three emerging bilinguals reported that it helped them visualize or focus.

**Item 8**

Shown in Figure 12 below, Item 8 assessed Life Sciences, and asked students about characteristics of mammals, using a platypus as an example. Although this content is taught in Grade 7, the three mainstream Grade 6 students and the one emerging bilingual Grade 6 student who took this item answered it correctly. Table 15 provides the item summary.

<table>
<thead>
<tr>
<th>A small animal called the duckbilled platypus lives in Australia. Which characteristic of this animal shows that it is a mammal?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. It eats other animals.</td>
</tr>
<tr>
<td>B. It feeds its young milk.</td>
</tr>
<tr>
<td>C. It makes a nest and lays eggs.</td>
</tr>
<tr>
<td>D. It has webbed feet.</td>
</tr>
</tbody>
</table>

*Figure 12. Item 8: Platypus*
Table 15

**Item 8: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>Illustrated</th>
<th>Not Illustrated</th>
<th>( n^1 )</th>
<th>Illustrated</th>
<th>Not Illustrated</th>
<th>( n^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answered correctly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Sound Out Science Term</td>
<td>3</td>
<td>- Sound Out Science Term</td>
<td>4</td>
<td>- Repeat Non-science Term</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>- Does not Understand Science Term</td>
<td>2</td>
<td>- Does not Understand Science Term</td>
<td>2</td>
<td>- Confusing</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>- Mispronounce Science Term</td>
<td>2</td>
<td>- Repeat Non-science Term</td>
<td>2</td>
<td>- Insert Non-science Term</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>- Mispronounce Science Term</td>
<td>2</td>
<td>- Mispronounce Science Term</td>
<td>2</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Problem Solving Strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Confirm Fact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Process of Elimination</td>
<td>1</td>
<td>- Partial Reasoning</td>
<td>2</td>
<td>- Recall Fact</td>
<td>4</td>
<td>- Process of Elimination</td>
</tr>
<tr>
<td>- Incorrectly Recall Fact</td>
<td>1</td>
<td>- Self Project</td>
<td>1</td>
<td>- Process of Elimination</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>- No Strategy</td>
<td>1</td>
<td>- Apply Fact</td>
<td>1</td>
<td>- Apply Fact</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>- Incorrectly Apply Fact</td>
<td>1</td>
<td>- Confirm Fact</td>
<td>1</td>
<td>- Confirm Fact</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- No Strategy</td>
<td>1</td>
<td>- Background Knowledge</td>
<td>1</td>
<td>- Background Knowledge</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Illustrate Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Did not Use</td>
<td>2</td>
<td></td>
<td></td>
<td>- Did not Use</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>- Focus on Aspect in Stem</td>
<td>1</td>
<td></td>
<td></td>
<td>- Better Understand</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- Make Connection to Everyday Experience</td>
<td>1</td>
<td></td>
<td></td>
<td>- Eliminate Answer</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- Interesting</td>
<td>1</td>
<td></td>
<td></td>
<td>- Interesting</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

\( ^1 \) Total \( n = 4 \)  
\( ^2 \) Total \( n = 5 \)
**Making sense of the item.** Emerging bilinguals and mainstream students greatly differed in the ways they made sense of this item. The terms *duckbilled platypus* and *characteristic* were terms that bilinguals sounded out, mispronounced, and reported they did not know across both illustrated and non-illustrated versions. No mainstream student was coded as sounding out, mispronouncing, or reporting that they did not know these terms. Surprisingly, two of the three emerging bilinguals who reported not knowing *duckbilled platypus* correctly answered the item by focusing on a word they did understand, *mammal*. Emerging bilinguals also substituted various non-science terms, including *lay* for *lays*, *character* for *characteristic*, and *eats* for *feeds*.

Mainstream students carried out very few actions to make sense of this item. In fact, for the item with an illustration, no code was utilized for more than one student to describe how they made sense of the item. In the non-illustrated version, monolinguals repeated non-science terms. Two mainstream reported this item was confusing. The first described the wording of option C as confusing, “it makes a nest and lays eggs.” The second mainstream student reported that the inclusion of extra information, that a platypus is from Australia, was confusing. Finally, two mainstream students inserted words when reading option B, “it feeds on its young milk” and “feeds its own young milk.”

**Problem solving strategies.** There was great heterogeneity in the strategies students used to solve the item. Emerging bilinguals who incorrectly answered this question often focused on the specific characteristics of a platypus rather than thinking about the entire class of mammals. Indeed, this item used an atypical example, a platypus, to exemplify the entire class of mammals. A Grade 6 emerging bilingual classified as LEP took the item without an illustration and was the only student, bilingual or monolingual, to note this exception:
S: I don’t think they [mammals] eat other animals but they might like kill them if they hurt their young, and I don’t actually know if they gave milk to their to their young, but I know it has webbed feet but they just use it to swim. And I know that they lay eggs ‘cause those are the only mammals that lay eggs.

The student was absolutely correct in all of his statements. However, he did not recognize that the item asked about all mammals, not just the platypus. He chose the incorrect option because he focused specifically on the platypus and not the entire class of mammals. Another emerging bilingual, a Grade 8 student classified as LEP, revealed that he disregarded the information about the platypus and only focused on what he knew about mammals to correctly solve the item. He stated,

R: What was the hardest part of this question?

S: Mm the when it said a small animal called a platypus ‘cause I didn’t really know what it was, and then I didn’t know that platypus was, like, a mammal, I just knew that mammals have their own milk.

The student successfully solved this problem by ignoring the extra information included in the stem. Because he focused on the most central aspect of the question (characteristics of mammals), he was able to select the right answer.

Not only does this item refer to an atypical example of a mammal, it forces students to compare an instance of a case to the entire class of cases. To carry out this comparison many students referred to other types of animals, both from the class of mammals and other classes of organisms. This was typically a successful strategy for students. An emerging bilingual classified as FEP took the item with no illustration and stated,
S: Well, I know that we’re mammals, and we don’t lay eggs like birds and we don’t make nests. We don’t—well, it does have webbed feet, but that doesn’t tell us anything that it’s a mammal or not. It eats other animals, yes, we do, too, but the basic one that tells you it’s a mammal is that it feeds its young milk, since we also feed young milk.

The student used herself, a mammal, and compared herself to other types of animals that are not mammals, birds. Then she narrowed down the answers to pick the correct choice. This was a common strategy also used by four mainstream students. A Grade 7 monolingual took the illustrated item and reported that distinguishing between the platypus’s characteristics and those of the general class of mammals was difficult:

R: What was the hardest part of this question?

S: Hmm, well, I just thought this one was a little bit difficult, because the platypus has a lot of different aspects that could be from different animal kingdoms, so, like, webbed feet, mammals don’t usually have webbed feet, and they don’t usually lay eggs. But they do feed their young with milk.

During her interview, the student showed that successful solvers must be able to keep a platypus’s characteristics separate from those of the entire class of mammals in order to answer correctly, something she found difficult.

**Use of vignette illustration.** Overall, the illustration did not appear to help students. Two of four emerging bilinguals and two of four mainstream students reported they did not use it. Those students who did use the illustration varied in their use. One emerging bilingual reported that she focused on the word *platypus* when she looked at the picture, while another reported that the illustration reminded her of a television show she saw. These uses differed from mainstream students who used reported using the illustration to remember what a platypus was, while
another used the illustration to eliminate option A, because she interpreted the plants around the platypus to mean that it was an herbivore and did not eat other animals.

**Summary of findings for Item 8.** Students from both linguistic backgrounds reported that this item contained unnecessary information. Rather than asking students about characteristics of mammals, the item used an atypical example that often confused students. Many emerging bilinguals reported they did not know or sounded out the words *duckbilled platypus*. In addition, this atypical example often encouraged students to focus on characteristics of the platypus rather than the class of mammals. The most successful strategy for students was to focus on characteristics of mammals by referring to other types of animals, including humans. Finally, half of students reported they did not use the illustration. Those students who did use the illustration used it for various reasons, again reflecting the heterogeneity in the ways that students interact with the item and the illustration.

**Item 9**

Item 9 assessed, Life Sciences, and asked students to describe how warm bloodedness helped mammals survive cold climates. It is shown in Figure 13 below.

> Which statement best explains why mammals are found in very cold regions of the world but lizards are not?

A. Both mammals and lizards are cold-blooded, but mammals have fur to keep them warm.
B. Both mammals and lizards are warm-blooded, but lizards get too cold when they shed their skin.
C. Since mammals, but not lizards, are warm blooded, their body temperature will adjust to match the external temperature.
D. Since mammals, but not lizards, are warm-blooded, they will maintain their body temperature.

*Figure 13. Item 9: Cold region*
This item was extremely difficult for emerging bilinguals. As Table 16 shows, no emerging bilingual answered correctly. It appeared fairly easy for mainstream students as three of five answered the illustrated version correct and three of four answered the non-illustrated version correctly. The content assessed with this item was taught in Grade 7. Two of the mainstream students who incorrectly answered were in Grade 6, and three emerging bilinguals who answered wrong were also in Grade 6, which could have explained why they answered incorrectly.

**Making sense of the item.** This item was difficult for emerging bilinguals to understand. They identified the following terms as unknown: *mammals, external, cold-blooded* and *warm-blooded*. Not knowing these terms was crucial, as each student who identified not knowing a science term incorrectly answered the item. In addition, two bilinguals described this item as confusing not only because of the science terminology, but the fact that the options were very similar in structure. Another difficult aspect to understand for emerging bilinguals were the logical connectors, such as *since, both, and but*. These words played an extremely important role in understanding the answer choices. This is explained by a Grade 7 student classified as FEP:

S: The hardest part was, like, ‘cause there’s some words that shake you. […] Like this one, it says, “both mammals and lizards,” and that’s both, so this is the same. And then “since mammals but not lizards,” it could confuse you there.

Even though this student understood the meaning of all these words, she found it confusing and difficult to keep these details straight among the options. This difficulty was echoed by a Grade 7 mainstream student taking the item with no illustration. She answered that the hardest part of the question was the similarity in the wording. She stated that, “It’s kind of all really similar.”

The hardest part for a Grade 6 bilingual classified as LEP, was to compare a specific type of animal (lizards) to a class of animals (mammals). He explained:
Table 16

**Item 9: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>ELLs Illustrated</th>
<th>ELLs Not Illustrated</th>
<th>Mainstream Illustrated</th>
<th>Mainstream Not Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered correctly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeat Non-science Term</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Substitute Science Term, Change Meaning</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Does not Understand Science Term</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Does not Understand Non-science Term</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitute Non-science term, Change Meaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confusing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeat Phrase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incorrectly Confirm Fact</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>Confirm Fact 4</td>
</tr>
<tr>
<td>Made Sense</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>Self Project 2</td>
</tr>
<tr>
<td>Everyday Experience</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Process of Elimination 2</td>
</tr>
<tr>
<td>Incorrectly Confirm Fact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incorrectly Recall Fact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illustration Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not Use</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connect to Everyday Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus on Aspect in Stem</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus on Aspect not in Stem</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Total n = 5
2 Total n = 4
S: They were comparing the mammals to lizards and they made the question harder like to have a good answer to it. […] Yah, the hard part was that they, like they compared mammals and lizards to it, two different kinds of animals.”

Rather than focusing their attention on one type of animal, students had to compare the characteristics of both types of animals and while deciphering the complex syntax structure. This was difficult for emerging bilinguals. A Grade 8 LEP student took the item with no illustration pointed this out:

S: Which statement best explains why mammals are found in very cold regions of the world but lizards are not. So it is just kind of confusing - why mammals are found in very cold regions of the world but lizards are not. […] So here I don’t get it because like I don’t really know what they are asking for . […] Statement best explains why mammals are found in very cold regions of the world but lizards are not. So, I’m thinking that lizards are not mammals or they are but they are not found on cold regions because they do not have fur to keep them warm so I would say it is B.

R: Can you recall your steps?

S: So, I read the question and then I read all these answers and then I read the question again and I just like did the one that made more sense to me.

R: Did any part help you solve it?

S: Well, no, not really cuz the question was confusing.

R: So, what was confusing about it?

S: Like, the very last part where it says like, very cold regions that mammals are not found in very cold regions of the world but lizards are not, yah so.
The linguistic features of this item were clearly challenging for this student. He was unable to decipher what the question was asking. During the follow up question he noted, that the statements ends with an ellipsis, *are not* which left the student to complete the sentence, *are not found*. During this bilingual’s verbalization he appeared to think that the question was stating that lizards were mammals, rather than comparing lizards to mammals.

**Problem solving strategies.** Many emerging bilinguals who answered incorrectly chose option A. These students typically focused on the part of the option that stated fur helped keep mammals warm. Although this is true, the first half of the answer choice stated that mammals are cold-blooded, a false statement. Unfortunately, many emerging bilinguals did not recognize this false statement and focused on the true information in the answer choice. A Grade 7 LEP student taking the item with no illustration justified his choice of A:

* S: Yah, because I think a lizard is a cold-blooded but it doesn’t have fur to keep it warm but mammals do, like bears and stuff.

This student provided a correct justification. Lizards are cold-blooded, and mammals have fur. Focusing only on these correct facts, he chose the distractor option.

In contrast to emerging bilinguals, the majority of mainstream students used multiple problem solving strategies for this item. Monolinguals who answered correctly were able to immediately eliminate the options stating that mammals were cold-blooded, and narrowed the options to C and D. Then, they focused on the fact that mammals were warm-blooded. Other mainstream students projected themselves onto the item, stating that humans were mammals. This was a successful strategy for only half of those students who invoked it. By self projecting herself onto the item, a Grade 7 monolingual stated:
S: And then this one, mammals are warm-blooded, because I know that I’m a mammal, and I don’t have fur on me all over, so that one isn’t right. And then this one, [pause] and then this one, I know that when I’m outside on a really hot day, my body temperature tries to cool itself, but that doesn’t always work, so it tries, but it doesn’t really adjust, it just gets hotter and hotter when it gets hotter and colder and colder when it gets cooler, so I just crossed that one out, and that’s the only one left.

By connecting aspects of herself, a warm-blooded mammal to the item, this monolingual selected the correct answer. No emerging bilingual used humans as an example of mammals.

**Use of vignette illustration.** Similar to previous items, some students reported not using this illustration, including two of five emerging bilinguals and three of five mainstream students. Those emerging bilinguals reported that they used the illustration, reported that they used it to focus on an aspect in the item. Unfortunately, this was not a successful strategy. One bilingual used the illustration to focus on the fur and stated, “I saw the picture and I saw the polar bear and I thought about how it helps keep it warm.” In this case, the illustration may have led the student to focus on the fur, something noted by a Grade 6 mainstream student who answered the question correctly:

S: Well, one thing that I thought actually kind of— [pause] um, one thing that I think wouldn’t have helped on this question is that it looked like the polar bear had fur, and that may have, um, made someone think of A, which is kind of not very helpful, because it’s not really the right answer.

This potential is further illustrated by a Grade 7 mainstream student who did choose option A. She reported that the illustration was helpful because “Well, like, it showed that there was fur and stuff on it, and that the polar bear was kind of in a cold climate and you didn’t see any
lizards there.” It appeared that the illustration led both a mainstream and an emerging bilingual to choose a distractor option. This again speaks to need for careful construction of vignette illustrations.

However, the illustration did help some students. A Grade 7 FEP student, answered the question wrong but reported that the illustration helped her set up the comparison in her mind. She described:

S: The image kind of helped me, because you could tell it was, like, a polar bear, or just a bear, and like, in the ice, you could tell, and since a polar bear is a mammal —

R: Can you say more about how that helped you?

S: Yeah, like, um, like, I think about polar bears and the temperature that they live in and comparing it to lizards, which is actually the question, lizards live in a hot climate, and if they’re both mammals or not.

Here, the illustrated appeared to help the bilingual focus on the fact that there was a mammal in a cold climate. Although the student answered this item incorrectly, as she was able to eliminate the other options after using the illustration set up the comparison in her mind.

**Summary of findings for Item 9.** This item was much more difficult for emerging bilinguals to understand than for mainstream students. Bilinguals identified not understanding science terms. In all cases, not understanding science terms resulted in choosing an incorrect answer option. In addition, some emerging bilinguals reported that they found the item confusing due to the complex syntax structure of the stem and the similar wording of the options.

Students used a variety of problem solving strategies. Students who successfully answered the item were able to eliminate options by knowing that mammals were warm-blooded.
Successful solvers were also able to relate warm-bloodedness to a mammal’s ability to maintain its body temperature.

The illustration was only helpful for a few students, as almost half of students (both emerging bilingual and mainstream) reported not using it. Emerging bilinguals who did use the illustration most frequently reported using it to focus on an aspect of the item. Unfortunately, in at least one case, the illustration led students to focus on the animal’s fur, and choose the incorrect answer option.

**Item 10**

Shown in Figure 14 below, the final item assessed Earth and Space sciences, a topic taught in Grade 8. As shown in Table 17, three of five students answered correctly across all comparative groups except for emerging bilinguals taking this item without an illustration, in which case only one student correctly answered.

If the temperature of Earth rose over time, which of the following would occur?

A. Sea level would fall, and the polar ice caps would decrease in size.
B. Sea level would fall, and the polar ice caps would increase in size.
C. Sea level would rise, and the polar ice caps would decrease in size.
D. Sea level would rise, and the polar ice caps would increase in size.

*Figure 14. Item 10: Earth’s Temperature*
Table 17

**Item 10: Frequencies of Most Common Codes by Linguistic Status and Coding Category**

<table>
<thead>
<tr>
<th>Code Type</th>
<th>Illustrated</th>
<th>Not Illustrated</th>
<th>Mainstream</th>
<th>Not Illustrated</th>
<th>n&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answered correctly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making Sense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Mispronounce Non-science Term</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>- Sound Out Science Term</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Omit Non-science Term</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Does not Understand Phrase</td>
<td>1</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>- Does not Understand Science Term</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Omit Non-science Term</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Apply Fact</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>- Self Project</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Guess</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Incorrectly Apply Fact</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Incorrectly Recall Fact</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- No Strategy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Process of Elimination</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Related Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Confirm Fact</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Process of Elimination</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Confirm Fact</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Process of Elimination</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Confirm Fact</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Background Knowledge</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Partial Reasoning</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Incorrectly Recall Fact</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Process of Elimination</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illustration Use</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Did not Use</td>
<td>4</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Confusing</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Heads Up</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup> Total n = 5
**Making sense of the item.** Students greatly varied in the ways they made sense of this item. Again, emerging bilinguals carried out more actions than mainstream students to make sense of this item. As shown in Table 17, seven emerging bilinguals mispronounced *occur*, but no student reported they didn’t know what this term meant. Compared to other items, there were much fewer instances of bilinguals reporting they did not understand science terms. Two emerging bilinguals reported they didn’t know what *polar ice caps* were. In both of these cases, students answered incorrectly. One other emerging bilingual reported not knowing *increase* and *decrease*, but still correctly solved the item.

For mainstream students taking the item with an illustration, there was no instance of more than one student for any given code to describe how they made sense of the item. This indicated that monolinguals read the item very close to the printed text, reported little difficulties understanding it, and did not use any observable strategies to make sense of the question. Similar to previous items, a frequent code for mainstream students taking the item with no illustration was to insert non-science terms including an article (the) and conjunctions (and).

Similar to the ways students made sense of Item 9, students reported that the similar wording of the options in this question made it confusing. Keeping track of the nuanced differences between the options was not easy for emerging bilinguals, as a Grade 8 LEP student pointed out:

*S:* I was confused, the answers they kind of gave all the same one, they kind of all say the same thing, and the the polar ice cap, and it changed right here, increase in size, decrease in size.

Even for students who understood all the words in this item, the nature of the answer choices made it difficult to fully understand each option.
**Problem solving strategies.** Students across all comparative groups greatly differed in the strategies they used to solve this item. One successful strategy that students from both linguistic groups used was to relate this item to the concept of global warming. One emerging bilingual and three mainstream students carried out this strategy. Interestingly, students only did this on the item with no illustration. Three of four students who carried out this strategy answered the question correctly. A Grade 7 student classified as FEP took the item with no illustration and related it to global warming during her think aloud:

S: Ok….so, creo que (silent) so creo que es C porque si sé que el calentamiento global está ocurriendo y toda la parte de Antártica y todas las partes donde vive el hielo están derritiendo, también que el nivel de los océanos está subiendo, so supe que es C porque decía que sea level would rise y sé que es real porque sé que el océano está subiendo y que también dice que polar ice caps would decrease in size que decrease significara como subiendo supe, eso sí como decrease como, I think, sí como, bajando, sí, eso. [Ok…so I think that (silent) so, I think that it is C because, yes I know that global warming is occurring and all the part of Antarctica and all the parts where there is ice is melting, also the ocean levels are rising and I knew it was C because it said that sea level would rise and I know that is real because I know the ocean is rising and it also says that the polar ice caps would decrease in size and that decrease means to rise, and I knew, yes, like decrease like, I think yes, like falling, yes that.]

Here, the bilingual student spoke, in Spanish, about her knowledge of global warming in general as well as the effect of global warming— that it causes the ice to melt and ocean levels to rise.

A successful strategy for many mainstream students, that was only utilized by one bilingual, was to confirm factual information provided in the stem. Four mainstream students (two on the
item with an illustration and two on an item with no illustration) correctly confirmed facts in this item, and answered the question right. In all four cases students confirmed that rising temperatures would cause ice to melt which would make sea levels rise and make the size of the ice caps decrease. Only one emerging bilingual attempted to confirm factual information in the question. He was incorrect in this confirmation and answered the item wrong.

**Use of vignette illustration.** There was a high level of similarity in the ways all students reported using this illustration. Unfortunately, the illustration did not appear to help students from either linguistic group. Four of five emerging bilinguels and three of five mainstream students reported not using this illustration. The one emerging bilingual who reported using the illustration seemed to be confused about what the illustration represented. Although he stated that it was a picture of a “temperature thing,” he also stated that it showed “sea level.”

Another mainstream student found the illustration confusing and stated that she didn’t know what the illustration was for. She says, “I don’t think it really fit in the question at all. That was just kind of strange, and it didn’t have any labels, and I didn’t know what that was for, so I just discarded it.” Only one mainstream student reported the illustration was helpful. He described that it, “it gives you, like, a heads-up of what it’s about.” Overall, this illustration did not appear to help either group of students.

**Summary of findings for Item 10.** Emerging bilinguals carried out more actions to make sense of this item than mainstream students. These actions were most frequently sounding out and mispronouncing science terms. Compared to other items, fewer bilinguals identified unknown science terms.

Similar to previous items, students used a wide variety of to solve this question. Successful strategies included relating the item to the concept of global warming and confirming factual
information provided in the item. The illustration for this item appeared to be unsuccessful. The majority of students reported they did not use this illustration while some found it confusing.

**Use of Spanish during Interviews**

A central goal during data collection was to provide bilingual participants with an inviting environment to use both English and Spanish. Although many students spoke with me in Spanish during casual interactions outside the interview setting, only eight bilinguals used Spanish during the interview.

There are several reasons that could explain why students may have chosen not to use Spanish during the interview. First, students attended an English only school, and were most likely not encouraged to communicate about academic topics in Spanish. In addition, the majority of emerging bilinguals learned about these topics in English and therefore, may have preferred to communicate about them using English. Finally, because I am a native English speaker, students could have been accommodating me by using English.

A previous analysis (Prosser & Solano-Flores, 2010) showed that students’ demographic background characteristics and English proficiency classifications did not predict which students prefer to use Spanish during the interview. That is, similar to the entire population of emerging bilinguals, the participants who used Spanish in this study come from a variety of backgrounds, had lived in the U. S. for varying amounts of time, started learning English at different ages, and were classified at different levels of English proficiencies. This heterogeneity is shown in Table 18.
The majority of students who preferred to use Spanish in this study had lived in the U.S. for 7 or more years. The variety in the background characteristics of participants using Spanish supports the sociolinguistic notion that all individuals choose to use language differently depending on the specific communicative event. I do not claim that these students were not able to express their reasoning using only English, but that they preferred to incorporate Spanish.

**Language use across interview sections.** The eight participants who used Spanish did so at different times throughout the interview, as shown in the Table 19 below.

Table 19

**Use of Spanish by Interview Section**

<table>
<thead>
<tr>
<th>Student</th>
<th>ELL status</th>
<th>Think aloud</th>
<th>Retrospective</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam$^1$</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delia</td>
<td>FEP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paz</td>
<td>FEP</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Andres</td>
<td>LEP</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Eduardo</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natalia</td>
<td>LEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alejandro</td>
<td>NEP</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Kati</td>
<td>LEP</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

$^1$All names used in this study are pseudonyms.
Two students only drew from Spanish during one section of the interview. Adam used Spanish only during his think aloud which is contrasted with Andrés and Natalia who drew from Spanish only during the follow-up questioning. These students differed from Paz, Alejandro, and Kati who all used Spanish in each of the three sections of the interview. The remaining four students used Spanish during two of the three interview sections. Although generalizing from this small number of participants is not possible, it is interesting to note that the two students classified at FEP preferred to communicate in Spanish during at least two sections of the interview. This indicates that one cannot assume students classified at various levels of proficiency will prefer to communicate in similar ways.

Recognizing that sociolinguistic principles state that the content of conversation influences language choice (Fishman, 1965), Table 20 compares use of Spanish across the interview sections for only the common items taken by all students.

Table 20

<table>
<thead>
<tr>
<th>Student</th>
<th>ELL</th>
<th>Concurrent</th>
<th>Retrospective</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>LEP</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delia</td>
<td>FEP</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paz</td>
<td>FEP</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Andrés</td>
<td>LEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eduardo</td>
<td>LEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natalia</td>
<td>LEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alejandro</td>
<td>NEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kati</td>
<td>LEP</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

The major difference for use of Spanish was that Andrés did not utilize Spanish when interacting with the common items. However, students’ language use remained fairly consistent when only analyzing Items 1 and 2. Some students classified at different levels of English proficiency used
Spanish during all three sections of the interview, such as Paz (FEP) and Kati (LEP). Other students only used Spanish during one or two sections (Adam, Natalia).

**Use of Spanish and other variables.** There appeared to be no relationship between a student’s use of Spanish and whether they correctly solved the item. Out of the twenty items in which students used Spanish when solving, eleven were answered correctly while nine were answered incorrectly. The results were similar for the relationship between use of Spanish and the presence of illustration as of the twenty items in which students used Spanish, eleven were illustrated and nine were not illustrated.

**Use of Spanish and problem solving strategies.** Bilingual students drew from Spanish when carrying out a variety of strategies. Some students used Spanish to recall things from the past, such as factual knowledge or classroom experiences. Other students utilized Spanish when applying facts or going through the process of elimination. Table 21 shows the strategies students carried out when speaking in Spanish. While there does not appear to be a major trend among these eight students, the most common use of Spanish across all items was with the code Apply Facts. Half of participants used Spanish to carry out this strategy. This second most frequent code was Process of Elimination which was a popular strategy for all students in the study. Table 22 provides the use of Spanish to strategies on Items 1 and 2 only, as different types of questions have been shown to elicit different types of reasoning (Ayala, Shavelson, Yin, & Schultz, 2002; Hamilton, Nussbaum, & Snow, 1997). Indeed, there were differences in students’ use of Spanish and their strategies on all items as compared to their use of Spanish and strategies on only the common items. There were a smaller number of strategies used across common items. However, the most popular codes are still Apply Fact and the Process of Elimination.
### Table 21

**Use of Spanish and Problem Solving Strategies**

<table>
<thead>
<tr>
<th>Student</th>
<th>ELL Status</th>
<th>Recall Fact</th>
<th>Background Knowledge</th>
<th>Class Experience</th>
<th>Everyday Experience</th>
<th>Apply Fact</th>
<th>Related Concept</th>
<th>Partial Reasoning</th>
<th>Process of Elimination</th>
<th>Guess</th>
<th>Self Monitor</th>
<th>Self Evaluate</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delia</td>
<td>FEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paz</td>
<td>FEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andres</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eduardo</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natalia</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alejandro</td>
<td>NEP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kati</td>
<td>LEP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 22

**Use of Spanish and Problem Solving Strategies: Common Items**

<table>
<thead>
<tr>
<th>Student</th>
<th>ELL</th>
<th>Recall Fact</th>
<th>Background Knowledge</th>
<th>Classroom Experience</th>
<th>Apply Fact</th>
<th>Partial Reasoning</th>
<th>Process of Elimination</th>
<th>Guess</th>
<th>Self Monitor</th>
<th>Self Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delia</td>
<td>FEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paz</td>
<td>FEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andres</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eduardo</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natalia</td>
<td>LEP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alejandro</td>
<td>NEP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kati</td>
<td>LEP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Purpose of using Spanish.** Finally, I investigated the purpose for which each student invoked Spanish. Three of the eight students used Spanish for lexical purposes. Rather than using Spanish more extensively, these three students only used Spanish to access one word, and then continued in English. Adam stated the following when thinking aloud to solve the Moon item. Classified as LEP, he thought out loud:

S: The moon produces no light and yet it shines at night. Why is this? The moon reflects the light from the sun, the moon rotates at very high speed, the moon is covered with a thin layer of ice, the moon have many cra-ters. I think it is A because the moon reflects the light from the sun because the sun is in the other side of the um earth and when it’s spinning, this, um *rayos del sol*, the moon catches those *rayos*. And then it reflects.

Here Adam substituted *rayos del sol* for sunrays. He quickly changed language to refer to one term, then returned to English to finish his verbalization. In a similar case, during the follow-up question, I asked another student, Andrés, what he found helpful in Item 7, Magnesium. He stated, “It helped me by, well the introduction helped me by like the oxy, like how do say that? Well, I said um, *por magnesio y oxígeno*, they both get, are magnesium oxide.” Here, Adam identified the Spanish cognates for magnesium and oxygen, a strategy of successful bilingual readers (Jiminez, 1996). Although he drew from Spanish to refer to only two words, he was able to continue verbalizing his reasoning by quickly switching to Spanish.

Five of the eight students used Spanish more extensively to describe their thinking, recall their steps, and answer follow up questions. The following excerpt is from Eduardo, a Grade 7 student classified as LEP, who had lived in the U.S. for eight years. He responded to my question, “Can you tell me the steps that you took to solve the question? ¿Los pasos que seguiste?”
S: Supe que los lizards no tienen. Lo leí todo y supe que era A porque lizards no tienen, son cold-blooded y no tienen fur y los otros mammals si y, so pueden vivir en the cold places but not lizards. [I knew that lizards don’t have any. I read everything and I knew it was A. because lizards don’t have any, they are cold-blooded and they don’t have fur, and other mammals do, so they can live in the cold places, but not lizards.]

By switching back and forth from English to Spanish he was able to explain his rationale for why he chose the answer he did. In this case, he focused the option’s true information, that mammals support cold temperatures because of their fur. He failed to recognize that the distractor option he chose incorrectly stated that mammals are cold-blooded.

However, by utilizing both English and Spanish he was able to fluidly explain his reasoning.

Paz, a Grade 7 student classified as FEP preferred to recall her steps entirely in Spanish.

S: Leí la pregunta y luego traté de ver lo que significaba y luego leí las opciones y luego traté de ver de lo que se trataba cada una y luego um pensaba que era C porque si derrite el hielo todo el agua se va para el océano y luego los ice caps se hacen más chiquitos.” [I read the question and then I tried to see what it meant and then I read the options and tried to see what each one was about, and I thought it was C. Because if the ice melts, all the water will go into the ocean and then the ice caps will be smaller.]

By utilizing Spanish, Paz recalled her steps and explained that ice melting would go into the ocean and would result in smaller ice caps. Similarly, Alejandro, a student classified
as NEP had been in the U.S. for two years, and was able to refer to the related concept (global warming) when choosing the correct answer. Perhaps, drawing from Spanish allowed him to refer to a concept he learned at his previous school in Mexico.

S: Ok….so, creo que (silent) so creo que es C porque si sé que el calentamiento global está ocurriendo y toda la parte de Antártica y todas las partes donde vive el hielo están derritiendo, también que el nivel de los océanos está subiendo, so supe que es C porque decía que sea level would rise y sé que es real porque sé que el océano está subiendo y que también dice que polar ice caps would decrease in size que decrease significara como subiendo supe, eso si como decrease como, I think, si como, bajando, si, eso. [Ok, so I think that (silent) so, I think that it is C because I know that global warming is happening throughout Antarctica and all the parts were the ice lives is melting, also, the level of the oceans is raising. So I knew it was C because it said, ‘sea level would rise’ and I know that is real because the ocean is rising and that also said that the ‘polar ice caps would decrease in size’ that decrease means to rise, I knew, that it’s like, I think, yes, lowering. Yes, this one.]

Here, Alejandro was able to apply what he knew about global warming to this item. Towards the end of his verbalization, he stated that decrease means to go up, but then corrected himself by stating it is actually to go down, making option C correct. Restricting Alejandro to English only during the interview may have not yielded as deep of insight about how he solved this item.

The results of how bilinguals chose to use language during the interviews speak to the heterogeneity of bilinguals’ language use in the same context. Although generalizing form these results is not possible due to the small number of students who chose to use
Spanish, they do highlight the varied roles a students’ first language plays while students think aloud and explain aspects of their reasoning.

**Summary of Results from Qualitative Analysis**

This chapter examined the heterogeneity in the ways students interacted with test items. An overall theme that emerged in the ways students made sense of items was that, for most items, emerging bilinguals carried out more actions than mainstream students. Many of these codes described how they deviated from the printed text when reading the item aloud. Additionally, emerging bilinguals more frequently identified terms they did not understand. The overwhelming majority of these terms were science vocabulary. Not understanding these terms was often associated with selecting the wrong answer. A few items were also challenging at the syntactical level, which made understanding the content of the item difficult for bilinguals.

Students greatly varied in the strategies they used to solve items. They varied across linguistic group, presence of illustration, and type of item. Each item appeared to have presented students with a unique situation. There was no overall trend observed in the way students solved each item. Typically, mainstream students carried out more problem solving strategies than emerging bilinguals. No one strategy was associated with correctly or incorrectly answering the item.

The use of illustration also varied by each item. For some items, the illustration was not successful, and many students reported not using the illustration or found it confusing. It is not surprising that many students reported not using the illustration, as the vignette illustration varies from other images in testing. Diagrams and charts are typically required to answer their corresponding test items and provide necessary information to solve it. Here, the vignette
illustration presents the same information found in the text of the item. Students may not be fully conscious of the nuanced role the illustration played when they solved the item.

A notable difference between the ways that bilinguals and monolinguals reported using the illustration was that bilinguals more frequently used the illustration to focus on an aspect of the item. This was not always associated with a correct answer, but it in some cases may have helped students better understand the content of the item.

A common thread in the analysis of bilinguals’ use of Spanish was the variety of ways students incorporated it throughout the interview. First, students used Spanish at different sections of the cognitive interview. Some students only used Spanish during one or two sections, whereas as other students used Spanish in all sections to think aloud, to recall their steps, and to answer follow-up questions. Second, much variation was observed in the strategies students carried out while using Spanish. Finally, some students invoked Spanish quite minimally, only speaking one term in Spanish whereas others used it more extensively to explain, justify, and identify the difficulty of various item features.

Test developers must not assume that language functions in the same way for students based on broad background characteristics or English proficiency classifications. If the field of assessment is serious about investigating issues of testing linguistically diverse students, they need to carry out detailed investigations to examine exactly how items function for each student. As shown in this chapter, students classified as fully proficient in English may prefer to express their reasoning in both English and Spanish. Educators and assessment developers must recognize these preferences and respect them when investigating how bilingual students interact with items. Restricting students to English only could constrict the information obtained and jeopardize the accuracy of conclusions drawn from students’ verbalizations.
Chapter 5
Quantitative Findings

This chapter discusses the quantitative results to investigate trends across linguistic groups and presence of illustration. I calculated frequency counts of codes to describe the ways students made sense of the items, their problem solving strategies, and their specific uses of the vignette illustration. For each research question, I first present the results of a two-way ANOVA to compare the statistical significance of differences in the frequency of codes at the cluster level. Second, radial graphs visually represent similarities and differences in the proportions of clusters of codes at the macro level. Third, I present symmetry graphs to investigate similarities and differences across linguistic groups and presence of illustration at the individual code level.

As noted in Chapter 3, the two common items (Items 1 and 2) were analyzed separately from the pool of randomly assigned items (Items 3-10). For the purpose of my analyses, Items 3-10 were assumed to be exchangeable and were collapsed due to the small number of students responding to each of these items. As a result, four sets of items were examined: Item 1 (illustrated), Item 2 (non-illustrated), Items 3-10 (illustrated versions), and Items 3-10 (non-illustrated versions). Analyzing Items 1 and 2 allows comparison across items given to all students. Analyzing Items 3-10 allows comparison with a larger sample of items.

Finally, I investigate the patterns in the percentages of students correctly answering items based on linguistic status and the presence of illustration. Next, a two-way repeated measures ANOVA shows the statistical significance in differences of student performance based on the presence of illustration and linguistic status.
Research Question 1

Making sense of the item at the cluster level. This section presents the quantitative results to answer the first research question: *How do emerging bilinguals and mainstream students compare in the ways that they make sense of multiple-choice science items, with and without illustrations?* Table 23 presents the results from a series of repeated measures two-way ANOVA analyses that shows the statistical significance of differences in the frequency of clusters of codes based on two factors: Linguistic Status (ELL or Non-ELL) and Presence of Illustration. The ANOVA results show no significant differences in the ways students made sense of items due to the interaction of the Presence of Illustration and Linguistic Status. Also, there are no statistically significant differences in the ways students made sense of items due to the main effect of Presence of Illustration. There are significant differences in the ways students made sense of items due to the main effect of Linguistic Status. ELLs were more frequently observed carrying out actions from the cluster of codes, Challenges to Understanding and Read Aloud Deviations. However, the effect size of these differences was small.

The radial graphs in Figures 15 and 16 present the relative frequencies (in percentages) for the clusters of codes across linguistic status and presence of illustration for Items 3-10. There are separate graphs to investigate if students differed when answering correctly or incorrectly. Figure 14 represents students who answered items correctly and Figure 15 represents students who answered the items incorrectly. Each point in the radial graph represents the proportion of total codes that a given cluster represents for that comparative group. The symmetric patterns of the dots in each quadrant of the graph indicate that students were similar in the ways they made sense of items. That is, there are similar proportions of each coding cluster across linguistic status and presence of illustration on correctly and incorrectly answered items.
Table 23

ANOVA Results for Frequency Differences of Codes for Making Sense by Linguistic Status and Presence of Illustration.

<table>
<thead>
<tr>
<th>Coding Cluster</th>
<th>Illustration</th>
<th>ELL</th>
<th>Illustration* ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>$\eta^2$</td>
</tr>
<tr>
<td>Challenges to Understanding</td>
<td>1.856</td>
<td>.177</td>
<td>.026</td>
</tr>
<tr>
<td>Strategies to Comprehend</td>
<td>.062</td>
<td>.805</td>
<td>.001</td>
</tr>
<tr>
<td>Read Aloud Deviations</td>
<td>.070</td>
<td>.793</td>
<td>.001</td>
</tr>
<tr>
<td>Corrections</td>
<td>.083</td>
<td>.775</td>
<td>.001</td>
</tr>
</tbody>
</table>
The clusters in the graphs above are taken from the major coding categories in the coding system described in Table 2 (Chapter 3). They include: Challenges to Understanding, Strategies to Comprehend, Read Aloud Corrections.
to Comprehend, Read Aloud Deviations, and Corrections to Read Aloud. The Challenges to Understanding cluster includes actions that described difficulties students had interpreting the item, such as identifying a science term they did not understand or identifying an aspect of the item as confusing. The Strategies to Comprehend cluster includes actions such as rereading, self-monitoring (in English or Spanish), and reformulating the item in one’s own words (in English or Spanish). Read Aloud Deviations includes actions such as repeating words, substituting words, and inserting words. Finally, the Corrections to Read Aloud category describes when students corrected any action from the Read Aloud Deviation, for example, correctly pronouncing a word after initially mispronouncing it.

The radial graph in Figure 15 shows emerging bilinguals making sense of illustrated items in the upper right quadrant. The highest proportion of codes for emerging bilinguals making sense of illustrated items is the Read Aloud Deviations cluster which made up approximately 75% of the total codes that describe how they made sense of items with an illustration. This finding is similar across linguistic groups and presence of illustration.

Figures 15 and 16 suggest an overall trend that the largest proportion of actions that students carried out to make sense of items was from the category, Read Aloud Deviations, followed by Challenges to Understanding, Strategies to Comprehend, and Corrections to Read Aloud. However, there are a few exceptions. First, as Figure 15 shows, for both bilinguals and monolinguals (the lower part of the graph), Challenges to Understanding on non-illustrated items make up the smallest proportion. This suggests that students who correctly answered non-illustrated items did not report as many difficulties understanding the items as students on the illustrated versions. Nevertheless, these graphs show a symmetric pattern and indicate that
students are more similar than different in the ways they make sense of items, regardless of linguistic background or presence of illustration.

**Items 3-10 at the individual code level.** To investigate differences and similarities of how students made sense of items at a finer grain level, Table 24 shows the frequency of the 10 most common codes for how students made sense of items (regardless of illustration). Table 24 provides contradictory information to the radial graphs as it shows both similarities and differences in the ways students made sense of items. Similarities across emerging bilinguals and monolinguals include the codes Sound Out Science Term, Substituting Non-science Term, Repeat Non-science Term, Does not Understand Science term, Mispronounce Science term, and Reread.

Table 24

*Making Sense: Frequencies of Most Common Codes, Regardless of Presence of Illustration*

<table>
<thead>
<tr>
<th>ELL</th>
<th>n</th>
<th>Non-ELL</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound Out Science Term</td>
<td>24</td>
<td>Insert Non-science Term</td>
<td>18</td>
</tr>
<tr>
<td>Substitute Non-science Term, Change Meaning</td>
<td>22</td>
<td>Omit Non-science Term</td>
<td>17</td>
</tr>
<tr>
<td>Repeat Non-science Term</td>
<td>20</td>
<td>Reread Item</td>
<td>16</td>
</tr>
<tr>
<td>Substitute Science term, Change Meaning</td>
<td>19</td>
<td>Repeat Non-science Term</td>
<td>14</td>
</tr>
<tr>
<td>Does not Understand Science Term</td>
<td>19</td>
<td>Confusing</td>
<td>13</td>
</tr>
<tr>
<td>Mispronounce Science Term</td>
<td>19</td>
<td>Substitute Non-science Term, Change Meaning</td>
<td>8</td>
</tr>
<tr>
<td>Reread Item</td>
<td>18</td>
<td>Sound Out Science Term</td>
<td>8</td>
</tr>
<tr>
<td>Substitute Non-science Term, Maintain Meaning</td>
<td>15</td>
<td>Repeat Science Term</td>
<td>8</td>
</tr>
<tr>
<td>Substitute Science Term, Maintain Meaning</td>
<td>15</td>
<td>Does not Understand Science Term</td>
<td>7</td>
</tr>
<tr>
<td>Sound Out Non-science Term</td>
<td>11</td>
<td>Mispronounce Science Term</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Repeat Phrase</td>
<td>7</td>
</tr>
</tbody>
</table>

However, there are also many differences in the ways students from different linguistic backgrounds made sense of items. First, there are a higher number of codes observed for ELLs
than for non-ELLs. This complements the ANOVA results and indicates that more emerging bilinguals carry out more actions to make sense of the item than mainstream students. Second, ELLs made more substitutions than non-ELLs, which included substituting science terms and non-science terms. Third, the two most frequent codes for mainstream students (Insert Non-science Term and Omit Non-science Term) are not among the most frequent codes for ELLs.

The symmetry graphs in Figures 17 and 18 compare ELLs and non-ELLs making sense of Items 3-10. The graphs rank the individual codes from most frequent (on the bottom) to least frequent (at the top). A symmetry graph shaped like a pyramid represents two similar groups. Symmetry graphs that do not resemble a pyramid shape represent two groups with different characteristics. Emerging bilinguals are on the left in blue and mainstream monolinguals are on the right in red. Figure 17 does not resemble a pyramid shape, indicating differences in the ways these students made sense of the items.

![Symmetry graph](image)

**Figure 17. Making sense: ELLs and non-ELLs**

Figure 18 below compares how all students (regardless of linguistic background) made sense of items with illustrations (in blue) to items with no illustrations (in red). This figure appears
more pyramid-like, indicating that there was a higher level of similarity in the ways students made sense of Items 3-10 with or without illustrations.

These two symmetry graphs complement the ANOVA results and indicate that there were more differences in the ways students made sense of items based on linguistic status (shown in Figure 17), than based on presence of illustration (shown in Figure 18).

To consider both linguistic status and presence of illustration, Table 25 lists the most frequent codes for how students made sense of items across linguistic group and presence of illustration. This table shows that ELLs were very similar in how they made sense of illustrated and non-illustrated items. On both types of items, the most frequent codes were Substitute Science and Non-Science Term, Sounded Out Science Term, and Not Understand Science Term. The symmetry graph in Figure 19 resembles a pyramid which further indicates similarities in the ways that emerging bilinguals made sense of illustrated and non-illustrated items.
### Table 25.

**Making Sense: Frequencies of Most Common Codes by Presence of Illustration and Linguistic Status**

<table>
<thead>
<tr>
<th>ELLs taking illustrated items</th>
<th>n</th>
<th>ELLs taking items without illustrations</th>
<th>n</th>
<th>Non-ELLs taking illustrated items</th>
<th>n</th>
<th>Non-ELLs taking items without illustration</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substitute Science Term, Change Meaning</td>
<td>11</td>
<td>Sound Out Science Term</td>
<td>15</td>
<td>Insert Non-science Term</td>
<td>9</td>
<td>Insert Non-science Term</td>
<td>9</td>
</tr>
<tr>
<td>Does not Understand Science Term</td>
<td>10</td>
<td>Substitute Non-science Term, Change Meaning</td>
<td>14</td>
<td>Omit Non-science Term</td>
<td>9</td>
<td>Confusing</td>
<td>9</td>
</tr>
<tr>
<td>Reread</td>
<td>9</td>
<td>Repeat Non-science Term</td>
<td>11</td>
<td>Reread</td>
<td>9</td>
<td>Repeat Non-science Term</td>
<td>9</td>
</tr>
<tr>
<td>Repeat Non-science Term</td>
<td>9</td>
<td>Mispronounce Science Term</td>
<td>11</td>
<td>Repeat Non-science Term</td>
<td>5</td>
<td>Omit Non-science Term</td>
<td>8</td>
</tr>
<tr>
<td>Sound Out Science Term</td>
<td>9</td>
<td>Does Not Understand Science Term</td>
<td>9</td>
<td>Repeat Phrase</td>
<td>5</td>
<td>Reread</td>
<td>7</td>
</tr>
<tr>
<td>Mispronounce Science Term</td>
<td>8</td>
<td>Reread</td>
<td>9</td>
<td>Substitute Non-science Term, Maintain Meaning</td>
<td>5</td>
<td>Sound Out Science Term</td>
<td>6</td>
</tr>
<tr>
<td>Mispronounce Non-Science Term</td>
<td>8</td>
<td>Substitute Science Term, Change Meaning</td>
<td>8</td>
<td>Correct Substitution of Non-science Term</td>
<td>5</td>
<td>Repeat Science Term</td>
<td>5</td>
</tr>
<tr>
<td>Substitute Non-science Term, Change Meaning</td>
<td>8</td>
<td>Substitute Science Term, Maintain Meaning</td>
<td>8</td>
<td>Confusing</td>
<td>4</td>
<td>Mispronounce Science Term</td>
<td>5</td>
</tr>
<tr>
<td>Substitute Science Term, Maintain Meaning</td>
<td>7</td>
<td>Substitute Non-science Term, Maintain Meaning</td>
<td>8</td>
<td>Does not Understand Science Term</td>
<td>4</td>
<td>Substitute Non-science Term, Change Meaning</td>
<td>4</td>
</tr>
<tr>
<td>Substitute Non-science Term, Maintain Meaning</td>
<td>7</td>
<td>Correct Substitution of Science Term</td>
<td>7</td>
<td>Substitute Non-science Term, Change Meaning</td>
<td>4</td>
<td>Does not Understand Science Term</td>
<td>3</td>
</tr>
<tr>
<td>Correct Insertion of Non-science Term</td>
<td>4</td>
<td>Omit Science Term</td>
<td>4</td>
<td>Correct Substitution Non-science Term</td>
<td>3</td>
<td>Correct Substitution Non-science Term</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 19. ELLs: Making sense by presence of illustration

Figure 20. Non-ELLs: Making sense by presence of illustration

Figure 21. Illustrated items: ELLs vs. non-ELLs

Figure 22. Non-illustrated items: ELLs vs. non-ELLs
Columns 3 and 4 of Table 25 compare mainstream students taking illustrated and non-illustrated items. Monolinguals were also very similar in the ways they made sense of illustrated and non-illustrated items. The most frequent codes for non-ELLs included Insert Non-science Term, Omit Non-science Term, and Reread. This high level of similarity is also shown by the pyramid shape in Figure 20.

Although students within the same linguistic status were similar in the ways they made sense of items, they greatly differed across linguistic status, shown by differences in the lists in Table 25. It is also shown by the non-pyramid-like shapes of Figures 21 and 22. Additionally, emerging bilinguals carried out more actions to make sense of the items than monolingual mainstream students, shown by the longer bars in blue in Figures 21 and 22. These graphs complement the statistically significant ANOVA results in the frequency of clusters of codes. These results indicate more differences in the ways students made sense of items due to linguistic status than based on the presence of illustration.

**Items 1 and 2 at the individual code level.** Table 26 lists the most frequent codes for how students made sense of Items 1 and 2. Similar to the results for Items 3-10, the frequencies of codes were higher for emerging bilinguals, indicating that they carried out more actions to make sense of the item than monolinguals. In making sense of Item 1, frequent codes for both groups of students were Substitute Non-science Term, Sound Out Science Term, and Reread.

Although Table 26 indicates similarities across ELLs and non-ELLs, the symmetry graph shows a slightly different story. As Figure 23 shows, there is only a moderate level of symmetry, as important differences are observed. The main difference is that the length of the bars are longer for emerging bilinguals. This length illustrates that emerging bilinguals carried out more actions to make sense of the item than mainstream students.
Table 26

Making Sense: Frequencies of Most Common Codes by Presence of Illustration and Linguistic Status: Items 1 and 2

<table>
<thead>
<tr>
<th>Item 1</th>
<th>ELL</th>
<th>Non-ELLs</th>
<th>Item 2</th>
<th>ELL</th>
<th>Non-ELLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>− Does not Understand Science Term</td>
<td>26</td>
<td>17</td>
<td>− Substitute Non-science Term, Change Meaning</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>− Sound Out Science Term</td>
<td>25</td>
<td>17</td>
<td>− Sound Out Science Term</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>− Substitute Non-Science Term, Change Meaning</td>
<td>23</td>
<td>14</td>
<td>− Substitute Science Term, Change Meaning</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>− Reread</td>
<td>22</td>
<td>8</td>
<td>− Substitute Non-science Term Maintain Meaning</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>− Mispronounce Science Term</td>
<td>20</td>
<td>7</td>
<td>− Correct Substitution of Non-science Term</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>− Substitute Science term, Change meaning</td>
<td>17</td>
<td>7</td>
<td>− Omit Non-science Term</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>− substitute Science Term, Maintain Meaning</td>
<td>13</td>
<td>6</td>
<td>− Substitute Science Term, Maintain Meaning</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>− Substitute Non-Science Term, Maintain Meaning</td>
<td>12</td>
<td>6</td>
<td>− Reread</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>− Sound Out Non-Science Term</td>
<td>12</td>
<td>4</td>
<td>− Repeat Non-science Term</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>− Correct Substitution of Non-science Term</td>
<td>11</td>
<td>4</td>
<td>− Correct Substitution of Science Term</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>− Confusing</td>
<td>4</td>
<td>4</td>
<td>− Confusing</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>− Repeat Science Term</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>


The two right most columns in Table 26 show the most frequent codes for how students made sense of Item 2 (without an illustration). These columns show many similarities in the ways all students made sense of Item 2, as many of the most frequent codes are the same across linguistic groups. These similarities include Substitute Non-science Term, Reread, and Confusing.

Although there are many similarities between emerging bilinguals and monolinguals, there are still many important differences, shown in Figure 24. First, there were far fewer mainstream students who carried out these actions to make sense of items. Second, many more mainstream students inserted a term into the text of Item 2 (an example, mentioned in the previous chapter, is the insertion of $a$, for the option that read, “The Moon rotates at $a$ very high speed”).

*Figure 23. ELLs vs. non-ELLs: Making sense of Item 1*
Research Question 1: More than One Instance of a Code

This section accounts for multiple instances of a code to compare how students made sense of items. For example, if a student identified not knowing what two science terms meant, it was recorded as a value of 2. By adding the number of times each code was used for each student, a total number was calculated that described how many actions students carried out to understand the item.

As explained in Chapter 3, the goal of calculating this number is to speak to previous research (e.g., Duran, 2008; Rivera, Collum, Wilner, & Sia, 2006) that has speculated that emerging bilinguals devote more cognitive resources to understanding items than mainstream students. While I do not argue that this number quantifies the cognitive resources a student utilized to make sense of the item, I believe it provides insight about the magnitude of differences between bilinguals and monolinguals when they make sense of items.

Table 27 shows the mean number of actions students carried out to make sense of items. This table shows that emerging bilinguals carried out more actions to make sense of all items, regardless of illustration. An independent samples t-test, shown in the column on the far right,
was statistically significant at the .000 level in all cases, and the effect sizes were medium to large. This indicates that there were major differences in the number of actions mainstream students and emerging bilinguals carried out to make sense of items.

Table 27

*Mean Number of Actions by Linguistic Status. Standard Deviations in Parentheses.*

<table>
<thead>
<tr>
<th></th>
<th>ELL</th>
<th>Non-ELL</th>
<th>Sig.</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Items 3-10</td>
<td>5.6</td>
<td>2.6</td>
<td>.000</td>
<td>.432</td>
</tr>
<tr>
<td></td>
<td>(3.9)</td>
<td>(2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items 3-10 illustrated</td>
<td>4.9</td>
<td>2.7</td>
<td>.000</td>
<td>.409</td>
</tr>
<tr>
<td></td>
<td>(2.9)</td>
<td>(2.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items 3-10 non-illustrated</td>
<td>6.3</td>
<td>2.4</td>
<td>.000</td>
<td>.484</td>
</tr>
<tr>
<td></td>
<td>(4.6)</td>
<td>(1.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1 illustrated</td>
<td>10.1</td>
<td>3.9</td>
<td>.000</td>
<td>.601</td>
</tr>
<tr>
<td></td>
<td>(5.2)</td>
<td>(2.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2 non-illustrated</td>
<td>4.3</td>
<td>1.4</td>
<td>.000</td>
<td>.527</td>
</tr>
<tr>
<td></td>
<td>(3.1)</td>
<td>(1.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For all types of items shown in Table 27, emerging bilinguals carried out a statistically significantly higher number of actions to make sense of items than mainstream students. However, an interesting difference appears when the items are separated by the presence of illustration. The mean number of actions for emerging bilinguals making sense of Items 3-10 in the illustrated versions was 4.9, and in the non-illustrated versions it was 6.3. That is, the mean was 1.4 actions lower with the presence of an illustration. In contrast, the mean number of codes for mainstream students remained constant across illustrated and non-illustrated items at 2.7 and 2.4 respectively. These results indicate that emerging bilinguals carried out fewer actions to make sense of illustrated items than for non-illustrated items.

To further investigate the number of actions students carried out to make sense of the items, Table 28 shows the mean number of actions by item and illustration. The first two rows of Table 28 show the mean scores for Item 1 (illustrated) and Item 2 (not illustrated). These two rows
show that the mean number of actions to make sense of the code was higher for the illustrated item. However, recall that these two items had very different linguistic features and should not be compared to each other.

Table 28 provides further evidence that emerging bilinguals carried out more actions than mainstream students to understand all items as the mean number of actions bilinguals carried out was higher in every case, except one. However, emerging bilinguals consistently carried out fewer actions to make sense of illustrated items than for non-illustrated items. The mean number of actions was higher for emerging bilinguals making sense of non-illustrated items than for illustrated items on Items 3-8 and Item 10. The mean number of actions to make sense of items for emerging bilinguals was higher only on illustrated Items 7 and 9. This indicates that, in general, emerging bilinguals carried out fewer actions to make sense of illustrated items. This trend was not observed for mainstream students.

The far right two columns in Table 28 show the differences between the number of actions mainstream students carried out and the number of actions emerging bilinguals carried out to make sense of illustrated or non-illustrated items. In all cases, except for Items 8 and 10, the difference in the number of actions students carried out was smaller for illustrated items. That is, the difference in the number of actions carried out to make sense of items between emerging bilinguals and mainstream students decreased on illustrated items.
Table 28

Mean Number of Actions by Illustration and Linguistic Group. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Item</th>
<th>Non-ELL Illustrated</th>
<th>Non-ELL Non-Illustrated</th>
<th>ELL Illustrated</th>
<th>ELL Non-Illustrated</th>
<th>Difference Between Groups Illustrated</th>
<th>Difference Between Groups Non-Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.9 (2.7) n=39</td>
<td></td>
<td>10.1</td>
<td>(5.2) n=39</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.4 (1.2) n=39</td>
<td></td>
<td>4.3</td>
<td>(3.1) n=39</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.1 (2.3) n=7</td>
<td>4.4 (2.3) n=5</td>
<td>6.0 (3.6)</td>
<td>9.8 (8.4)</td>
<td>2.9</td>
<td>5.4</td>
</tr>
<tr>
<td>4</td>
<td>3.5 (2.4) n=4</td>
<td>2.0 (2.0) n=5</td>
<td>5.0 (3.8)</td>
<td>8.4 (6.0)</td>
<td>1.5</td>
<td>6.4</td>
</tr>
<tr>
<td>5</td>
<td>1.6 (1.5) n=5</td>
<td>1.6 (1.7) n=7</td>
<td>3.4 (1.5)</td>
<td>8.0 (7.0)</td>
<td>1.8</td>
<td>6.4</td>
</tr>
<tr>
<td>6</td>
<td>2.8 (1.3) n=5</td>
<td>3.6 (1.1) n=5</td>
<td>6.2 (3.9)</td>
<td>6.6 (1.7)</td>
<td>2.4</td>
<td>3.0</td>
</tr>
<tr>
<td>7</td>
<td>5.0 (4.0) n=5</td>
<td>4.6 (2.1) n=5</td>
<td>6.0 (2.3)</td>
<td>7.0 (2.6)</td>
<td>1.0</td>
<td>2.4</td>
</tr>
<tr>
<td>8</td>
<td>1.0 (0.7) n=5</td>
<td>2.0 (1.2) n=4</td>
<td>5.0 (1.8)</td>
<td>4.3 (2.3)</td>
<td>4.0</td>
<td>2.3</td>
</tr>
<tr>
<td>9</td>
<td>4.2 (1.9) n=5</td>
<td>1.4 (0.9) n=4</td>
<td>3.8 (2.6)</td>
<td>5.6 (3.0)</td>
<td>-0.4</td>
<td>4.2</td>
</tr>
<tr>
<td>10</td>
<td>1.2 (0.8) n=5</td>
<td>3.4 (3.4) n=5</td>
<td>6.4 (2.7)</td>
<td>3.8 (1.6)</td>
<td>5.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Next, Table 29 shows the mean number of actions students carried out to make sense of items across grade levels. It does not indicate any trend. There appears to be no relationship between the number of actions students carried out to make sense of items and their grade level.
Table 29

Mean Number of Actions to Make Sense of Items by Grade and Linguistic Status. Standard Deviations in Parentheses.

<table>
<thead>
<tr>
<th></th>
<th>ELL</th>
<th>Non-ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 7 8</td>
<td>6 7 8</td>
</tr>
<tr>
<td>Illustrated 3-10</td>
<td>5.4 4.5 6.0</td>
<td>2.5 3.1 2.3</td>
</tr>
<tr>
<td></td>
<td>(3.4) (2.6) (3.1)</td>
<td>(2.0) (2.7) (1.5)</td>
</tr>
<tr>
<td>Not Illustrated 3-10</td>
<td>4.6 5.3 10.1</td>
<td>2.5 2.9 1.7</td>
</tr>
<tr>
<td></td>
<td>(3.1) (3.0) (6.6)</td>
<td>(2.1) (2.2) (1.2)</td>
</tr>
<tr>
<td>Illustrated 1</td>
<td>11.5 8.2 13.3</td>
<td>4.1 3.5 4.6</td>
</tr>
<tr>
<td></td>
<td>(5.6) (4.5) (4.7)</td>
<td>(2.3) (3.2) (2.4)</td>
</tr>
<tr>
<td>Not Illustrated 2</td>
<td>5.2 4.6 3.4</td>
<td>1.7 1.2 1.3</td>
</tr>
<tr>
<td></td>
<td>(4.0) (3.4) (1.4)</td>
<td>(1.5) (1.0) (1.3)</td>
</tr>
</tbody>
</table>

Table 30 examines the mean number of actions emerging bilinguals carried out to make sense of items by their classification of English proficiency. There are only three students classified as NEP, which requires caution when interpreting the results. Again, there was no overall trend observed and it appears that there is no relationship between the number of actions emerging bilinguals carried out to make sense of items and their classification of English proficiency.

Table 30

Mean Number of Actions to Make Sense of Items by ELL Classification.

<table>
<thead>
<tr>
<th></th>
<th>NEP</th>
<th>LEP</th>
<th>FEP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 3</td>
<td>n = 24</td>
<td>n = 9</td>
</tr>
<tr>
<td>Illustrated Item 1</td>
<td>8.7</td>
<td>10.7</td>
<td>8.6</td>
</tr>
<tr>
<td>Not Illustrated Item 2</td>
<td>6.0</td>
<td>4.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Illustrated 3-10</td>
<td>5.7</td>
<td>5.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Not Illustrated 3-10</td>
<td>8.3</td>
<td>7.3</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Table 30 above shows that students classified as FEP typically carried out the least number of actions to make sense of the item. However, students classified as LEP carried out more
actions than students classified as NEP on Item 1 and illustrated Items 3-10. Therefore, there appears to be no trend between English proficiency classification and the number of actions carried out.

**Summary of findings for Research Question 1.** This section showed that emerging bilinguals carried out significantly more actions to make sense of items. Specifically, ANOVA results showed they were observed as carrying out more actions from the Challenges to Understanding and Read Aloud Deviation clusters of codes. However, the proportion of codes to describe how ELLs and non-ELLs made sense items was similar, as shown by symmetry in the radial graphs. Nonetheless, analyzing students’ actions at a finer grain level revealed important differences, shown in symmetry graphs of individual codes. These symmetry graphs also highlighted the fact that emerging bilinguals carried out more actions to make sense of items than mainstream students.

The results on the total number of actions students carried out to make sense of the item provided further evidence that ELLs carried out more actions to make sense of items than Non-ELLs. This difference was statistically significant across all items, with and without illustrations. However, when items were analyzed individually, ELLs carried out fewer actions to make sense of illustrated items than non-illustrated items. This was not the case for mainstream students as there was no trend in the number of actions they carried out to make sense of the item based on illustration. Analyzing the number of actions emerging bilinguals carried out to make sense of the item revealed no trend based on grade level or classification of English proficiency.

**Research Question 2**

This section presents the quantitative results to answer the second research question: *How do emerging bilinguals and mainstream students compare in the strategies they use to solve*
multiple-choice science items, with and without vignette illustrations? First, two way ANOVAs show the statistical significance in the frequency of codes across linguistic status and presence of ELLs. Second, radial graphs explore trends across linguistic groups and presence of illustration at the code cluster level. These clusters are: (1) Background Knowledge (Recall Facts, Recall Classroom Experiences, Recall Everyday Experiences, and Apply Facts), (2) In the Moment Knowledge (Confirm Fact and Self Project), and (3) Testing Strategies (Process of Elimination, Self-Monitor, Guess, or No Strategy). Third, symmetry graphs investigate similarities and differences at the individual code level.

**Problem solving strategies at the cluster level.** Table 31 presents the results for the two-way ANOVA in which Presence of Illustration and Linguistic Status were treated respectively as within- and between subject-factors to determine the statistical significance of differences in the frequency of codes for problem solving strategies. The ANOVA results show that there are no significant differences in the strategies students used to solve items based on the interaction of Presence of Illustration and Linguistic Status. The main effect of Presence of Illustration is also not significant. There are significant differences in the strategies they used to solve based on Linguistic Status. Non-ELLS used significantly more strategies from the cluster of codes, In the Moment Knowledge, than ELLs. However, a small effect size indicates these differences are most likely not noticeable as shown in the radial graphs in Figures 25 and 26. The symmetric pattern of these graphs indicates that there were more similarities than differences in the strategies students used to solve items across linguistic groups and presence of illustration on Items 3-10. For almost all the comparative groups, the largest proportion of codes was Testing Strategies, followed by Background Knowledge, and In the Moment Knowledge. This trend is observed across items answered correctly and incorrectly.
Table 31

*ANOVA Results for Differences in the Frequencies of Codes for Problem Solving Strategies*

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Illustration F</th>
<th>Illustration Sig.</th>
<th>Illustration $\eta^2$</th>
<th>Linguistic Status F</th>
<th>Linguistic Status Sig.</th>
<th>Linguistic Status $\eta^2$</th>
<th>Illustration* Linguistic Status F</th>
<th>Illustration* Linguistic Status Sig.</th>
<th>Illustration* Linguistic Status $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>.070</td>
<td>.792</td>
<td>.001</td>
<td>2.466</td>
<td>.121</td>
<td>.034</td>
<td>.279</td>
<td>.599</td>
<td>.004</td>
</tr>
<tr>
<td>In the Moment Knowledge</td>
<td>.382</td>
<td>.538</td>
<td>.006</td>
<td>4.678</td>
<td>.034</td>
<td>.063</td>
<td>.382</td>
<td>.538</td>
<td>.006</td>
</tr>
<tr>
<td>Testing Strategies</td>
<td>.072</td>
<td>.789</td>
<td>.001</td>
<td>.182</td>
<td>.671</td>
<td>.003</td>
<td>.000</td>
<td>1.000</td>
<td>.000</td>
</tr>
</tbody>
</table>
Figure 25. Problem solving strategies: Correctly answered Items 3-10.

Figure 26. Problem solving strategies: Incorrectly answered Items 3-10.

**Problem solving strategies at the individual code level.** This section compares emerging bilinguals and mainstream monolinguals in their problem solving strategies at the individual
code level. Process of Elimination was the most frequent code for students from both linguistic
groups across illustrated and non-illustrated items. This is not surprising as the structure of
multiple-choice items presents students with many options from which they have to choose one.
Other than this commonality, there were important differences in students’ problem solving
strategies. As shown in Table 32 below, the most frequent strategies for emerging bilinguals
were No Strategy, Made Sense, and Partial Reasoning. This differed from the most frequent
strategies for mainstream students which included Recall Facts, Confirm Facts, and Apply Facts.

Table 32

Frequencies of Most Common Codes for Problem Solving Strategies: Items 3-10, Regardless of
Presence of Illustration

<table>
<thead>
<tr>
<th></th>
<th>ELLs</th>
<th>n</th>
<th>Non-ELLs</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>− Process of Elimination</td>
<td>12</td>
<td></td>
<td>− Process of Elimination</td>
<td>25</td>
</tr>
<tr>
<td>− No Strategy</td>
<td>11</td>
<td></td>
<td>− Recall Facts</td>
<td>13</td>
</tr>
<tr>
<td>− Made Sense</td>
<td>8</td>
<td></td>
<td>− Confirm Facts</td>
<td>12</td>
</tr>
<tr>
<td>− Partial Reasoning</td>
<td>8</td>
<td></td>
<td>− Apply Facts</td>
<td>9</td>
</tr>
<tr>
<td>− Guess</td>
<td>8</td>
<td></td>
<td>− Self Monitor</td>
<td>9</td>
</tr>
<tr>
<td>− Incorrectly Confirm Fact</td>
<td>6</td>
<td></td>
<td>− Focus</td>
<td>8</td>
</tr>
<tr>
<td>− Classroom Experience</td>
<td>5</td>
<td></td>
<td>− Made Sense</td>
<td>7</td>
</tr>
<tr>
<td>− Incorrectly Apply Fact</td>
<td>4</td>
<td></td>
<td>− Self Projection</td>
<td>6</td>
</tr>
<tr>
<td>− Recall Facts</td>
<td>4</td>
<td></td>
<td>− Partial Reasoning</td>
<td>6</td>
</tr>
<tr>
<td>− Apply Facts</td>
<td>4</td>
<td></td>
<td>− Self Evaluate</td>
<td>5</td>
</tr>
<tr>
<td>− Self Monitor</td>
<td>3</td>
<td></td>
<td>− Guess</td>
<td>4</td>
</tr>
<tr>
<td>− Focus</td>
<td>3</td>
<td></td>
<td>− Related Concept</td>
<td>3</td>
</tr>
<tr>
<td>− Self Projection</td>
<td>3</td>
<td></td>
<td>− Incorrect Confirmation of Fact</td>
<td>3</td>
</tr>
<tr>
<td>− Self Evaluate</td>
<td>2</td>
<td></td>
<td>− Classroom Experience</td>
<td>3</td>
</tr>
<tr>
<td>− Related Concept</td>
<td>2</td>
<td></td>
<td>− General Background Knowledge</td>
<td>2</td>
</tr>
<tr>
<td>− Incorrectly Recall of Fact</td>
<td>2</td>
<td></td>
<td>− No Strategy</td>
<td>1</td>
</tr>
<tr>
<td>− Everyday Experience</td>
<td>1</td>
<td></td>
<td>− Incorrectly Apply Facts</td>
<td>1</td>
</tr>
<tr>
<td>− Confirm Fact</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 27 provides the symmetry graph that compares ELLs and non-ELLs solving all items
(regardless of the presence of illustration). It confirms that there were some similarities but also
some important differences in the problem solving strategies emerging bilinguals and mainstream students used. While there is a moderate pyramid shape, mainstream students carried out more problem solving strategies than emerging bilinguals, indicated by the length of the bars in the graph.

![Bar chart showing problem solving strategies for ELLs and Non-ELLs](image)

*Figure 27. Problem solving strategies: ELLs vs. non-ELLs*

Figure 28, shown below, compares all students (regardless of linguistic status) taking illustrated and non-illustrated items. There is a high level of symmetry represented by the pyramid-shaped figure in the graph. This shape indicates that there were more similarities in the problem solving strategies across illustrated and non-illustrated items than across linguistic groups. That is, there were more differences between emerging bilinguals and mainstream monolinguals than between illustrated and non-illustrated items.
To explore the similarities and differences across linguistic groups and presence of illustration, Table 33 lists the most frequent strategies for each comparative group. Process of Elimination was the most frequent strategy for all cases, except for emerging bilinguals taking illustrated items. Emerging bilinguals were mostly similar in the strategies they used across items with and without illustrations. On both types of items there were many emerging bilinguals that were coded as, No Strategy or Partial Reasoning. The few differences that existed between emerging bilinguals included the codes Made Sense and Recall Classroom Experience which were more frequent on illustrated items than non-illustrated items. The existence of both similarities and differences are shown in Figure 29 below. The graph follows a pyramid shape, and indicates that there are some similarities but also some differences in the ways that emerging bilinguals solve items with or without the accommodation.

Next, the third and fourth columns in Table 33 list the strategies mainstream students used to solve items. There were more similarities than differences in the strategies that monolinguals used to solve items with and without illustrations. The most frequent codes were Recall Fact,
Table 33

**Frequencies of Most Common Codes for Problem Solving Strategies by Linguistic Status and Presence of Illustration**

<table>
<thead>
<tr>
<th>ELLs taking illustrated items</th>
<th>n</th>
<th>ELLs taking items without illustrations</th>
<th>n</th>
<th>Non-ELLS taking illustrated items</th>
<th>n</th>
<th>Non-ELLS taking items without illustration</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Strategy</td>
<td>7</td>
<td>Partial Reasoning</td>
<td>5</td>
<td>Recall Facts</td>
<td>7</td>
<td>Confirm Fact</td>
<td>8</td>
</tr>
<tr>
<td>Process of Elimination</td>
<td>6</td>
<td>Guess</td>
<td>5</td>
<td>Apply Facts</td>
<td>7</td>
<td>Recall Fact</td>
<td>6</td>
</tr>
<tr>
<td>Incorrectly Recall Facts</td>
<td>4</td>
<td>No Strategy</td>
<td>4</td>
<td>Confirm Facts</td>
<td>4</td>
<td>Self Monitor</td>
<td>6</td>
</tr>
<tr>
<td>Classroom Experience</td>
<td>4</td>
<td>Incorrectly Apply Fact</td>
<td>3</td>
<td>Focus</td>
<td>4</td>
<td>Focus</td>
<td>5</td>
</tr>
<tr>
<td>Apply Facts</td>
<td>3</td>
<td>Recall Facts</td>
<td>3</td>
<td>Made Sense</td>
<td>4</td>
<td>Apply Fact</td>
<td>4</td>
</tr>
<tr>
<td>Partial Reasoning</td>
<td>3</td>
<td>Focus</td>
<td>3</td>
<td>Self Evaluate</td>
<td>4</td>
<td>Made Sense</td>
<td>4</td>
</tr>
<tr>
<td>Guess</td>
<td>3</td>
<td>Incorrectly Confirm Fact</td>
<td>2</td>
<td>General Background Knowledge</td>
<td>3</td>
<td>Self Project</td>
<td>4</td>
</tr>
<tr>
<td>Incorrectly Apply Fact</td>
<td>2</td>
<td>Self Monitor</td>
<td>2</td>
<td>Self Project</td>
<td>3</td>
<td>Partial Reasoning</td>
<td>4</td>
</tr>
<tr>
<td>Self Projection</td>
<td>2</td>
<td>Self Evaluate</td>
<td>2</td>
<td>Self Monitor</td>
<td>3</td>
<td>Related Concept</td>
<td>4</td>
</tr>
<tr>
<td>Self Monitor</td>
<td>2</td>
<td>Incorrectly Recall Fact</td>
<td>2</td>
<td>Partial Reasoning</td>
<td>2</td>
<td>Guess</td>
<td>3</td>
</tr>
<tr>
<td>Recall Fact</td>
<td>1</td>
<td>Made Sense</td>
<td>1</td>
<td>Graphic Fact</td>
<td>2</td>
<td>Incorrectly Confirm Fact</td>
<td>3</td>
</tr>
<tr>
<td>Confirm Fact</td>
<td>1</td>
<td>Classroom Experience</td>
<td>1</td>
<td>No Strategy</td>
<td>2</td>
<td>Classroom Experience</td>
<td>3</td>
</tr>
<tr>
<td>Everyday Experience</td>
<td>1</td>
<td>Apply Facts</td>
<td>1</td>
<td>Incorrectly Confirm Fact</td>
<td>1</td>
<td>Self Evaluate</td>
<td>2</td>
</tr>
<tr>
<td>Related Concept</td>
<td>1</td>
<td>Self Projection</td>
<td>1</td>
<td>Classroom Experience</td>
<td>1</td>
<td>Incorrect Recall of Fact</td>
<td>1</td>
</tr>
<tr>
<td>Self Evaluate</td>
<td>1</td>
<td>Everyday Experience</td>
<td>1</td>
<td>Incorrectly Apply Fact</td>
<td>1</td>
<td>Everyday Experience</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Related Concept</td>
<td>1</td>
<td>Visualize</td>
<td>1</td>
<td>Incorrectly Apply Fact</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 29. Problem solving strategies: ELLs

Figure 30. Problem solving strategies: Non-ELLSs

Figure 31. Problem solving strategies: Illustrated items

Figure 32. Problem solving strategies: Non-illustrated items
Apply Fact, Confirm Facts, and Focus on Aspect. The moderate pyramid shape of Figure 29 above confirms these similarities.

There were many differences in the problem solving strategies of emerging bilinguals and monolinguals. These differences are shown in the most frequent strategies listed in Table 33. Figures 31 and 32 also indicate that were more differences than similarities between these two groups of students, as the graphs do not resemble a strong pyramid shape.

To investigate trends across students on the common items, Table 34 lists the strategies each group of students used to solve Items 1 and 2. Again, there were both similarities and differences in the ways students solved these items. Note that the number of mainstream students was typically larger than the number of emerging bilinguals carrying out a given strategy. This indicates that mainstream students carried out more problem solving strategies than emerging bilinguals.

When solving Item 1, for both groups of students, frequent codes included Process of Elimination and Recall Fact. However, a more frequent code for emerging bilinguals than for mainstream students was Guess or No Strategy. The symmetry graph in Figure 33 below indicates that there were some similarities across emerging bilinguals and mainstream students solving Item 1. However, there were differences, such as no observations of the codes Guess or No Strategy for mainstream students.
Table 34

*Items 1 and 2: Frequencies of Most Common Codes for Problem Solving Strategies*

<table>
<thead>
<tr>
<th>Item 1</th>
<th>ELL</th>
<th>Mainstream</th>
<th>n</th>
<th>Item 2</th>
<th>ELL</th>
<th>Mainstream</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall Fact</td>
<td>9</td>
<td>Process of Elimination</td>
<td>16</td>
<td>No Strategy</td>
<td>7</td>
<td>Process of Elimination</td>
<td>17</td>
</tr>
<tr>
<td>Process of Elimination</td>
<td>8</td>
<td>Recall Fact</td>
<td>15</td>
<td>Background Knowledge</td>
<td>5</td>
<td>Apply Fact</td>
<td>14</td>
</tr>
<tr>
<td>Guess</td>
<td>6</td>
<td>Made Sense</td>
<td>7</td>
<td>Apply Fact</td>
<td>5</td>
<td>Classroom Experience</td>
<td>8</td>
</tr>
<tr>
<td>No Strategy</td>
<td>5</td>
<td>General Background Knowledge</td>
<td>5</td>
<td>Process of Elimination</td>
<td>5</td>
<td>General Background Knowledge</td>
<td>5</td>
</tr>
<tr>
<td>Made Sense</td>
<td>3</td>
<td>Apply Fact</td>
<td>5</td>
<td>Recall Fact</td>
<td>4</td>
<td>Recall Fact</td>
<td>5</td>
</tr>
<tr>
<td>Self Monitor</td>
<td>3</td>
<td>Self Monitor</td>
<td>2</td>
<td>Confirm Fact</td>
<td>4</td>
<td>Everyday Experience</td>
<td>4</td>
</tr>
<tr>
<td>Incorrectly Confirm Fact</td>
<td>2</td>
<td>Classroom Experience</td>
<td>2</td>
<td>Classroom Experience</td>
<td>4</td>
<td>Made Sense</td>
<td>4</td>
</tr>
<tr>
<td>Classroom Experience</td>
<td>2</td>
<td>Incorrectly Apply Fact</td>
<td>2</td>
<td>Incorrectly Apply Fact</td>
<td>4</td>
<td>Partial Reasoning</td>
<td>4</td>
</tr>
<tr>
<td>Incorrectly Apply Fact</td>
<td>2</td>
<td>Partial Reasoning</td>
<td>2</td>
<td>Everyday Experience</td>
<td>3</td>
<td>Incorrectly Recall Fact</td>
<td>2</td>
</tr>
<tr>
<td>Partial Reasoning</td>
<td>2</td>
<td>Incorrectly Recall Fact</td>
<td>2</td>
<td>Made Sense</td>
<td>3</td>
<td>Self Evaluate</td>
<td>2</td>
</tr>
<tr>
<td>Self Evaluate</td>
<td>2</td>
<td>Incorrectly Confirm Fact</td>
<td>1</td>
<td>Incorrectly Recall Fact</td>
<td>2</td>
<td>Incorrectly Confirm Fact</td>
<td>2</td>
</tr>
<tr>
<td>Incorrectly Recall Fact</td>
<td>1</td>
<td>Self Evaluate</td>
<td>1</td>
<td>Focus</td>
<td>2</td>
<td>Self Project</td>
<td>2</td>
</tr>
<tr>
<td>Confirm Fact</td>
<td>1</td>
<td>Related Concept</td>
<td>1</td>
<td>Self Evaluate</td>
<td>2</td>
<td>Confirm Fact</td>
<td>1</td>
</tr>
<tr>
<td>General Background Knowledge</td>
<td>1</td>
<td>Focus</td>
<td>1</td>
<td>Incorrectly Confirm Fact</td>
<td>1</td>
<td>Guess</td>
<td>1</td>
</tr>
<tr>
<td>Related Concept</td>
<td>1</td>
<td>Visualize</td>
<td>1</td>
<td>Guess</td>
<td>1</td>
<td>Real Object</td>
<td>1</td>
</tr>
<tr>
<td>Focus</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For Item 2, there were many similarities in the strategies that emerging bilinguals and mainstream students used to solve items. The biggest difference was that No Strategy was the most frequent code for emerging bilinguals, but not for mainstream students. Process of Elimination was again a popular strategy for students from both linguistic groups, but more mainstream students (17) used this approach than emerging bilinguals (5). Other similarities included the codes, Apply Fact, Recall Fact, and Classroom Experience.

The symmetry graph in Figure 34 below illustrates that mainstream students and emerging bilinguals used similar problem solving strategies for Item 2. However, the length of the bars showed that more mainstream students used these strategies.
Overall, these symmetry graphs show that there were many differences in the problem solving strategies of students from different linguistic backgrounds. However, there were some cases, such as Item 2, where monolingual and bilingual students used similar strategies. These similarities and difference suggests that there was an interaction between student, item, and illustration.

**Summary of findings for Research Question 2.** This section showed both similarities and differences in the ways students solved items. This was shown at both the cluster level and individual code level. One major similarity between ELL and mainstream students was that Process of Elimination was used by students from both linguistic groups in the presence and absence of illustrations. One major difference was that a frequent code for emerging bilinguals was No Strategy or Partial Reasoning whereas a frequent code for mainstream student was Recall Fact and Confirm Fact.

**Research Question 3**

**Use of vignette illustration at the cluster level.** This section answers the final research question: *How do emerging bilinguals and mainstream monolinguals compare in their specific*
use of the vignette illustration? Table 35 lists the most frequent codes used to describe how students used the vignette illustration. The left two columns of the table compare the uses of illustrations for Items 3-10 and the right two columns of the table list the uses for Item 1. Both emerging bilinguals and mainstream students frequently reported not using the illustration. This is not surprising as vignette illustrations are different from the traditional images found on science tests including diagrams, graphs, and other illustrations which students must use to arrive at an answer. It is possible students were not aware of the nuanced role the illustration played in their thought processes.

Other similarities in the ways students reported using the illustrations included using them to focus on an aspect of the item. A code that was more frequent for emerging bilinguals was Focus on Aspect of Stem. In contrast, a more frequent code for mainstream students was Focus on Entire Stem or Focus on Aspect not Mentioned in Stem. Additionally, mainstream students more frequently identified the illustrations as confusing, represented by code Confusing.

Figure 35 provides the symmetry graph for illustration use which indicates both similarities and differences in the ways students used illustrations on Items 3-10.

Figure 35. Vignette illustration use: Items 3-10
Table 35

*Frequencies of Most Common Codes for Use of Vignette Illustration by Linguistic Group and Item*

<table>
<thead>
<tr>
<th>Items 3-10</th>
<th>ELL</th>
<th>Mainstream</th>
<th>n</th>
<th>ELL</th>
<th>Mainstream</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>− Did not Use</td>
<td>14</td>
<td>− Did not Use</td>
<td>17</td>
<td>− Support Answer Choice</td>
<td>9</td>
<td>− Better Understand Item</td>
</tr>
<tr>
<td>− Focus on Aspect in Stem</td>
<td>7</td>
<td>− Confusing</td>
<td>5</td>
<td>− Did not Use</td>
<td>7</td>
<td>− Support Answer Choice</td>
</tr>
<tr>
<td>− Visualize Aspect in Stem</td>
<td>4</td>
<td>− Focus on Entire Stem</td>
<td>3</td>
<td>− Focus on Aspect in Stem</td>
<td>7</td>
<td>− Did not Use</td>
</tr>
<tr>
<td>− Focus on Entire Stem</td>
<td>2</td>
<td>− Focus on Aspect not in Stem</td>
<td>2</td>
<td>− Focus on Aspect not in Stem</td>
<td>4</td>
<td>− Visualize Aspect in Stem</td>
</tr>
<tr>
<td>− Focus on Aspect not in Stem</td>
<td>2</td>
<td>− Better Understand</td>
<td>2</td>
<td>− Connect to Science Topic</td>
<td>3</td>
<td>− Interesting</td>
</tr>
<tr>
<td>− Connect to Everyday Experience</td>
<td>2</td>
<td>− Focus on Aspect in Stem</td>
<td>1</td>
<td>− Better Understand Item</td>
<td>2</td>
<td>− Focus on Aspect in Stem</td>
</tr>
<tr>
<td>− Connect to Classroom Experience</td>
<td>2</td>
<td>− Connect to Everyday Experience</td>
<td>1</td>
<td>− Confusing</td>
<td>2</td>
<td>− Confusing</td>
</tr>
<tr>
<td>− Confusing</td>
<td>1</td>
<td>− Support Answer Choice</td>
<td>1</td>
<td>− Focus on Entire Stem</td>
<td>2</td>
<td>− Visualize Entire Stem</td>
</tr>
<tr>
<td>− Support Answer Choice</td>
<td>1</td>
<td>− Eliminate Answer Choice</td>
<td>1</td>
<td>− Visualize Aspect in Stem</td>
<td>1</td>
<td>− Connect to Science Topic</td>
</tr>
<tr>
<td>− Eliminate Answer Choice</td>
<td>1</td>
<td>− Connect to Science Topic</td>
<td>1</td>
<td>− Visualize Aspect not in Stem</td>
<td>1</td>
<td>− Focus on Entire Stem</td>
</tr>
<tr>
<td>− Connect to Background Knowledge</td>
<td>1</td>
<td>− Connect to Classroom</td>
<td>1</td>
<td>− Look for Answer</td>
<td>1</td>
<td>− Visualize Aspect not in Stem</td>
</tr>
<tr>
<td>− Visualize Entire Stem</td>
<td>1</td>
<td>− Give Heads Up</td>
<td>1</td>
<td>− Connect to Background Knowledge</td>
<td>1</td>
<td>− Eliminate Answer Choice</td>
</tr>
<tr>
<td>− Visualize Aspect not in Stem</td>
<td>1</td>
<td>− Interesting</td>
<td>1</td>
<td>− Gives Heads Up</td>
<td>1</td>
<td>− Gives Heads Up</td>
</tr>
</tbody>
</table>

192
Next, the right two columns in Table 35 show the most frequent illustration use for Item 1. Similarities included the codes Support an Answer Choice and Did not Use. One difference between emerging bilinguals and mainstream students was that mainstream monolinguals more frequently coded as Better Understand, indicating the illustration helped them better comprehend the item. Although this is surprising, the qualitative analysis described how monolinguals often reported the illustration clarified details not provided in the text of the item.

Another difference between students was that bilinguals frequently reported using the illustration to focus on aspects of Item 1. The symmetry graph in Figure 36 below indicates both similarities and differences in the ways ELLs reported using the illustration for this item.

![Figure 36. Vignette illustration use: Item 1](image)

**Summary of findings Research Question 3.** These results showed both similarities and differences in vignette illustration use, as shown in the symmetry graphs. However, there were important differences, including that emerging bilinguals more frequently used the illustration to focus on an aspect of the item. These results provide further evidence for an interaction between the student, item, and illustration.
Student Performance on Illustrated and Non-illustrated Versions of Items

**Percent of students correctly answering items.** To investigate if the impact of illustrations on student performance, I calculated the percentages of students answering correctly by linguistic status and presence of illustration. If illustrations do provide the intended support for students to gain access to the content of items, more students should respond correctly to illustrated items than items without an illustration.

Table 36 shows the results of student performance based on presence of illustration and correctness of answer by each item. For each item, the results are presented for emerging bilinguals, mainstream students, and then for all students at the far right of the table. This table showed no trend in the percentage of students answering correctly across the presence of illustration and correctness of answer. Based on these results, it cannot be confirmed or disconfirmed that illustrations result in students answering the item correctly. Rather, due to the heterogeneity in the percentages of students answering correctly across linguistic groups, items, and presence of illustration, these results provide evidence for an interaction between these three factors.

**Performance based on linguistic status and presence of illustration.** To investigate statistically significant differences between students from different linguistic groups solving items with and without illustration, I performed a repeated measures, Linguistic Status x Presence of Illustration, two-factor ANOVA. Linguistic status was treated as a between-subjects factor and presence of illustration was treated as a within-subjects factor. The results in Table 37 below show that there are no statistically significant differences in student performance based on the interaction of Presence of Illustration and Linguistic Status. Similarly, there are no significant differences based on the main effect of Presence of Illustration. There are significant differences
Table 36

*Student Performance across Illustrated and Non-illustrated Versions of Items: Rounded Percentages*

<table>
<thead>
<tr>
<th>Item</th>
<th>ELL Illustrated</th>
<th>Not Illustrated</th>
<th>Non-ELL Illustrated</th>
<th>Not Illustrated</th>
<th>All Students Illustrated</th>
<th>Not Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>25</td>
<td>13</td>
<td>50</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>33</td>
<td>22</td>
<td>22</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>38</td>
<td>0</td>
<td>38</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>43</td>
<td>0</td>
<td>29</td>
<td>29</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>50</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>60</td>
<td>0</td>
<td>40</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 37

*ANOVA Results for Student Performance by Linguistic Status and Presence of Illustration*

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>Sig.</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustration</td>
<td>.654</td>
<td>.421</td>
<td>.009</td>
</tr>
<tr>
<td>Linguistic status</td>
<td>14.078</td>
<td>.000</td>
<td>.167</td>
</tr>
<tr>
<td>Illustration* Linguistic status</td>
<td>.000</td>
<td>1.000</td>
<td>.000</td>
</tr>
</tbody>
</table>
based on the factor, Linguistic Status, although the effect size is small. These results show that non-ELLs tended to correctly answer items more than ELLs.

Summary of Results from Quantitative Analysis

This chapter presented the quantitative results for all three research questions. Two-way ANOVAs showed statistically significant differences in the frequency of codes for how students made sense of the item at the cluster level based on linguistic status. These clusters included Overall Comprehension and Read Aloud Deviations. Radial graphs showed that, at the cluster level, students were similar in the ways they made sense of items. However, analysis at the individual code level revealed important differences. Overall, there were more similarities within linguistic groups than between linguistic groups. The presence of illustration did not appear to influence how either group of students made sense of the item. It was clear that emerging bilinguals carried out many more actions to make sense of items than mainstream monolingual students. The number of actions emerging bilinguals carried out was significantly higher than the number of actions mainstream students carried out. However, the mean number of actions to make sense of items was slightly lower for emerging bilinguals on illustrated items than non-illustrated items.

Students from different linguistic backgrounds varied in the types of strategies they used to solve items. This was shown in the statistically significant results of the ANOVA for the frequency of codes from the cluster of codes, In the Moment Knowledge based on linguistic status. Radial graphs indicated that students were similar in the proportion of strategies they carried out from each of the cluster of codes. However, symmetry graphs at the individual code level revealed important differences in the problem solving strategies of ELLs and non-ELLs.
Many students from both linguistic backgrounds reported not using the vignette illustration. Perhaps this was because the illustration was not required to solve the item and students may not have been conscious of the specific way they utilized the illustration. Nevertheless, there were differences in the ways emerging bilinguals and mainstream students reported using the illustration. These differences included emerging bilinguals using the illustration to focus on an aspect of the item.

Investigating the percentage of students who answered correctly based on linguistic status and presence of illustration revealed no trend. Rather, it suggests an interaction between student, item, and illustration. Repeated measures ANOVA showed no significant differences based on the interaction of the presence of illustration and linguistic status. There were also no significant differences based on the main effect of presence of illustration. However, the ANOVA showed that non-ELLS scored statistically significantly higher the ELLs.
Chapter 6

Discussion

The purpose of this study was to provide empirical evidence, based on verbal protocols, on the similarities and differences between Spanish-speaking emerging bilinguals and native English speaking middle school students solving multiple-choice science items with and without vignette illustrations, as a form of testing accommodation. The analysis focused on three issues: how students made sense of the items, their problem solving strategies, and their specific uses of the vignette illustration. After discussing the results, this chapter will examine the results as they relate to large-scale assessment of emerging bilinguals, classroom teaching and assessment of emerging bilinguals, and the ITELL study. Additionally, I discuss the study’s limitations and make suggestions for future research.

Assessing emerging bilinguals is an extremely complex endeavor. The entanglement of language and content often leads to inaccurate results of what linguistically diverse students know (Solano-Flores, 2006; 2008). Historically, test development has paid little attention to emerging bilinguals throughout the phases of test development. This lack of consideration is extremely problematic, especially when the tests are used to make decisions about students, teachers, and schools. This study highlights important differences in the ways emerging bilinguals and mainstream students interact with test items. This interaction is dependent on each emerging bilingual’s set of strengths and weaknesses and the unique linguistic challenges posed by each item. The result is a dynamic interaction between the student and the item’s linguistic features. Due to the variety of ways that students will interact with different test items, assessment systems must involve a large number of emerging bilinguals into their practices of developing items. Kopriva (2000; 2001) has argued that the field needs more in-depth qualitative
research and more studies utilizing cognitive interviews with emerging bilinguals in testing situations. This study showed that valuable insight is obtained when this advice is followed.

Not only is there a great need for more research in the testing of emerging bilinguals, this research must be carried out using appropriate theoretical support (Valdés & Figueroa, 1994). The scant research that does investigate issues of assessing emerging bilinguals often utilizes a monolingual framework (Grosjean, 1985; 1989). This framework leads to an incomplete picture of how emerging bilinguals interact with test items. By drawing from theories of holistic bilingualism, second language acquisition, and sociolinguistics, this study was able to gain deeper insight into the ways students solve multiple-choice science items. It serves as an example for the assessment community of ways that these theories can be utilized to develop more accurate measures of what linguistically diverse students know.

Qualitative analysis illuminated important differences between monolinguals and bilinguals in their experiences responding to test items. Each item provided students with a unique set of challenges, of which they handled in different ways. An overall trend was seen that emerging bilinguals carried out more actions to make sense of the items than mainstream monolingual students. These differences included that many more emerging bilinguals reported not understanding science vocabulary terms than mainstream students. While some students were able to focus on the words in the items they did understand to correctly solve the item, this was not always a successful strategy. The ability to understand science terms was often essential to correctly solving the problem. However, no trend was observed between not understanding a science term and grade level, English proficiency classification, or another factor.

While previous research shows that ELLs typically need 5-7 years to obtain English proficiency (Hakuta, Butler, & Witt, 2000), this study aligns with Collier’s (1995) claim that it
can take longer than seven years for emerging bilinguals to become proficient in aspects of academic English and the scientific register. Fourteen bilingual participants were born in the U.S. and the average number of years living in the U.S. for those students born outside the U.S. was 7 years. Though details of students’ previous schooling are unknown, it is true that they attended English only middle schools, and most likely attended English only elementary schools. The fact that, after 7 years in the U.S., ELLs frequently reported not understanding science terms strongly questions the effectiveness of English only programs. It is likely that, due to the fact that students were not proficient in English, they did not have the same opportunity to learn the science content as their mainstream counterparts (see Abedi, 2007; Solórzano, 2008). This could have played a role in ELLs’ unfamiliarity with science terms.

Although I encouraged all bilinguals to draw from both languages when solving items, the majority of bilinguals participated in the cognitive interview in only English. The fact that the majority of students did not use Spanish most likely reflects the fact that the participants have not been encouraged or allowed to use their first language in academic settings. Therefore, while there is a push to view students’ native language as a resource (Ruiz, 1984), language will not necessarily be a resource if students have been educated in an English only system. Such systems do not provide the opportunity to learn content in their native language, nor does it provide opportunities to strategically draw from the native language. For most of the students in this study, Spanish was most likely viewed as a social language and not considered a resource for academic tasks. A more effective system would view Spanish as a resource that students can strategically and purposefully use to help them accomplish academic tasks (see Baker, 2008; Hopewell, 2011).
However, for some students, Spanish was a resource when solving items. Nevertheless, students’ use of Spanish greatly varied, and was not associated with grade, topic, or classification of English proficiency. Test developers must not use general background characteristics to make assumptions about the ways students will use language during assessment situations (Solano-Flores, 2009). Appropriately addressing the needs of linguistically diverse students requires detailed investigations of how each item functions across many bilingual students, allowing them to draw from both languages to express how they interact with test items.

Overall, students’ problem solving strategies greatly varied across and within linguistic groups. That is, for any given item, multiple different strategies were used by emerging bilinguals and monolingual students. However, a trend emerged that mainstream monolinguals carried out more strategies than emerging bilinguals, often utilizing a combination of strategies. This provides further evidence that students from different linguistic groups solve items in different ways. The most frequent code observed for emerging bilinguals was that they did not use any strategy. Although it is possible that these results were a result of bilinguals’ verbalization style during the interview, they speak to the need for the inclusion of linguistically diverse students in the process of test development as the strategies intended by the item writers may not coincide with the strategies students invoke.

Traditionally, accommodations are the main tool used to include students in testing. However, these accommodations are developed as an afterthought, with little acknowledgement for the needs of linguistically diverse (Solano-Flores, 2010a). The vignette illustrations were designed drawing from theories of bilingualism and second language acquisition to better fit the needs of emerging bilinguals. Unfortunately, many illustrations examined in this study did not appear to help students. Perhaps this was due to the contents of the illustration or its features.
Future research is needed to determine how these illustrations can really help ELLs in testing situations. The results from this study showed that students’ uses of the vignette illustration varied across linguistic groups and item, indicating an interaction between student, item, and illustration. While the majority of all students reported not using the illustration, it helped many emerging bilinguals focus on an aspect of some of the items. In contrast, mainstream students often reported the illustration helped them better understand details of the question. However, it was clear that, for a few items, neither group of students benefited from the presence of the vignette illustration. Future research should focus on developing procedures for creating illustrations that can maximize the potential for the illustration to help students understand the item.

A final important finding was the difference in quantitative results depending on the level of analysis. ANOVA results indicated statistically significant results in the frequency of clusters of codes including Read Aloud Deviation, Challenges to Understanding, and In the Moment Knowledge. Radial graphs showed that, at the level of coding cluster, students from different linguistic groups appeared to be similar in the ways they made sense of items with and without the vignette illustration. When this analysis was disaggregated to the individual code level, important differences were observed. Accurate interpretation of results required analysis at a fine grain level, qualitatively and quantitatively analyzing individual codes. Focusing purely on quantitative analysis at the macro level would not have revealed important differences between emerging bilinguals and mainstream monolinguals. Researchers and assessment developers must consider this when investigating aspects of how students interact with their test items.
Implications

Implications for classroom teaching and assessment. Teachers need to know the needs of individual students. Results from this study show that multiple-choice items may not provide accurate information about what emerging bilinguals know. In many cases, students showed they understood the concept being assessed but chose a distractor. Solely interpreting a student’s answer choice as right or wrong discounts what he/she may know. Although time consuming, think aloud interviews provided detailed information about why each student chose a specific answer. Teachers should work to include think aloud or think aloud-type activities in their instruction to gain more comprehensive input about what their students know. This information should inform their teaching about what aspects of a given topic need to be emphasized or re-taught.

The results of this study showed that various science terms were not well understood by many emerging bilinguals. Although previous research showed that everyday polysemic words can be problematic in the testing of emerging bilinguals (e.g., Abedi, Hofstetter, Lord, 2004; Sato, Rabinowitz, Gallagher, & Huang, 2010), in this study, science terms with only one meaning were the most frequent terms not understood. Terms such as seedling, photosynthesis, exert, and duckbilled platypus were all identified by multiple emerging bilinguals as unknown. This calls for science content teachers to focus on the language of science in their teachings. Although, the students’ science teachers were seasoned educators in their field, none of them were bilingual and they did not hold degrees in the field of English as a second language. Perhaps instruction that incorporates both of students’ languages strategically would be more effective to learning the content and language of science. At the very least, professional development in English as a Second Language methodology should be required for content
teachers so that their daily activities are effective for all students, both monolingual and bilingual.

Even though vocabulary was highlighted in this study as an area for growth, students also identified not understanding phrases, or discourse features used in the entire stem of the item. Teachers need to explicitly teach the vocabulary and the language structures of science on a daily basis. This instruction should go beyond simply defining terms. Research has shown that linguistically diverse students need multiple contacts across differing contexts to learn technical vocabulary (see Herrera, Perez, & Escamilla, 2010). Successful science instruction necessarily includes a focus on the language of the discipline.

Finally, teachers should recognize that emerging bilinguals may need more time to make sense of academic tasks. Linguistically diverse students should be allowed sufficient time to complete classroom tasks and assessments. As the results of this study showed, students all levels of English proficiency classification can benefit from extra time. Therefore, sufficient time on academic tasks should be given to ELLs at all classifications.

**Implications for large-scale assessment.** The results of this study echo what some researchers (e.g., Abedi, Hofstetter, & Lord, 2004; Solano-Flores, 2009; Valdés & Figueroa, 1994) have argued. Important differences exist in the ways that emerging bilinguals and mainstream monolinguals make sense of test items. This study extended previous studies by providing detailed evidence from student think alouds about these differences. Based on these differences, there is an urgent need to include emerging bilinguals in all phases of test development. This inclusion must be carried out utilizing appropriate theoretical support, which views linguistically diverse students as holistic bilinguals and considers both languages.
The major finding of this study is that emerging bilinguals clearly carry out more actions to make sense of test items. This fact and the fact that they utilized different strategies than those used by monolinguals have important implications for the use of test results. Concluding that emerging bilinguals do not know the content assessed based on their selection of a distractor is problematic because, as this study highlighted, students often know the content being assessed, but the linguistic features of the item do not allow them to demonstrate this knowledge.

Another implication for assessment systems is the need to pilot test items with a large number of linguistically diverse students. Due to diversity in the strategies that emerging bilinguals carried out when solving items, test developers should increase the number of bilinguals they include in their piloting phases. Piloting items with only a few students will not provide comprehensive insight into the variety of strategies that emerging bilinguals carry out to solve items. This piloting should be an iterative process in which the items are refined and re-piloted with emerging bilinguals to further investigate how the items function.

Due to the results that showed emerging bilinguals carried out more actions to make sense of items than mainstream students, extra time is an appropriate accommodation in all testing situations. Moreover, these results showed that emerging bilinguals did not necessarily vary in the number of actions they carried out to make sense of items based on their classification of English proficiency. That is, students classified at all levels of English proficiency carried out many actions to make sense of items. Therefore, serious consideration needs to be given to the types of accommodations allowed to students at all levels of English proficiency. A student who is reclassified at higher levels of proficiency may still benefit from extra time to participate in standardized tests.
Finally, this study provided an example of how theories of bilingualism and second language acquisition can be utilized to more appropriately investigate how emerging bilinguals solve test items with and without accommodations. Large-scale testing programs should take note and incorporate these theories into their practices. As Abella, Urrutia, and Schneyderman (2005) state, “at present time, there exists no specific criterion for determining when standardized achievement tests can validly assess students’ knowledge bases when tested in a second language” (p. 128). Assessment systems need to seriously consider ways that they can improve their practices to obtain more accurate measures of what linguistically diverse students know.

**Implications for the ITELL project.** This study showed that the vignette illustration has potential to help students better understand and focus on aspects of the item. This potential should be investigated through an iterative process that informs the development of the illustrations. Although there were no significant differences in the performance of students based on the presence of an illustration, it is important to recall that the illustrations used in this investigation were not developed with the final version of the procedure. More research is needed to understand the exact nature of how the illustration may help or hinder emerging bilinguals’ access to the content of the item. Because vignette illustrations differ from traditional illustrations due to the fact that they are not required to answer the test question, students may not have been aware of how the illustration influenced their interaction with test items. Further studies are needed to continue to determine the exact role of vignette illustrations in solving test items. This future research should include multiple different types of illustrations and test items as the current study suggests an interaction between student, item, and illustration. In other words, an illustration can be effective in supporting ELLs to gain access to the content of items for some students depending on the item and the characteristics of the illustration.
The results of this study speak to the need for the ITELL study to pilot all vignette illustrations during their development. Some illustrations, as seen with the Item 1, Seedling, may lead students to choose a specific answer choice. Other illustrations may not be effective for bilinguals or monolinguals, and may even be distracting or confusing, as was the case with Item 7, Magnesium. Because the illustrations functioned differently across students, this piloting should include a large number of participants.

Limitations

One limitation of this study is self-selection. Attempts to gain consent forms from all Spanish-speaking emerging bilinguals were not entirely successful. However, the participants in this study represent the majority of bilinguals at this school, as 36 out of 43 total emerging bilinguals agreed to participate. The comparative group, monolinguals, also self selected into this study with a much lower rate of response, at about 30% of the entire population of monolinguals at this school. Therefore, the random sample of monolinguals selected to participate in this study may not be completely representative of the entire population of mainstream students at this school.

A second limitation of this study is the inability to generalize to other schools and other populations of emerging bilinguals. The school from which all participants attended had specific characteristics that most likely differ from other schools in the district, state, and nation. It was an English-only school with a relatively small population of emerging bilinguals. Therefore, it is unknown the extent to which these results can be generalized to other emerging bilinguals attending bilingual schools and/or schools with larger populations of emerging bilinguals. In addition, all the emerging bilinguals in this study spoke Spanish as a native language and the results do not necessarily reflect emerging bilinguals who speak other languages.
Due to the small number of emerging bilinguals at this school, the number of bilingual participants in Grade 8 was small. Therefore, the participant pooled extended to include students from Grades 6 and 7, which resulted in some students answering test items that covered content in a grade level they had yet to reach. It does not appear that this played a major role in the results, as often students from the lower grades correctly answered items assessing content of the upper grades. Nonetheless, results may have differed had the study only included participants only from Grade 8.

Next, the results of this study were dependent on student verbalizations. These verbalizations reflect the metacognitive abilities of students. For example, students need to know that they do not understand a word to report whether they comprehend it. Furthermore, due to communication styles, student verbalizations may have differed with another interviewer, perhaps one who was more familiar to the students or who was a native-Spanish speaker. Regardless of interviewer characteristics, student verbalizations may not reflect the exact nature of how students interacted with test items. However, triangulating results across the various interview sections provided the most accurate insight possible.

Finally, this study did not use independent coding. While the consensus coding resulted in solid coding, results may have differed had independent coding been utilized. Given the exploratory nature of this study, consensus coding was the best approach to coding the data.

**Future Research**

During this study, several aspects arose as questions to guide further research. First, the fact that emerging bilinguals carried out more observable actions to make sense of the text brought up questions of automaticity as it relates to students’ experience solving test items. In literacy, automaticity is an important aspect of fluent reading and is defined as “quick, effortless
identification of words that enables the reader to focus his or her attention on the cognitive task of comprehending instead of decoding” (Herrera, Perez, & Escamilla, 2010). As previously mentioned, scholars have speculated that emerging bilinguals devote more cognitive resources to understanding items than their mainstream monolingual peers. While this speculation is not necessarily debated, the results of this study led me to ponder how test items can be written so emerging bilinguals can more fluently read them and focus their attention on the content assessed. Linguistic modification studies have not been successful in minimizing the role of language on emerging bilinguals’ performance (see Keiffer, Lesaux, Rivera, Francis, 2009). However, the methodology used to investigate the detailed interaction of student and item in this study could provide new insight about what linguistic features are most problematic for emerging bilinguals solving multiple-choice items. Future research should examine the role of various linguistic aspects identified in this coding system to determine which codes are most important. For example, what really happens when students repeat a word? Perhaps this word was unfamiliar to them, requiring them to devote resources to recalling or figuring out its meaning. Lord, Abedi, and Poosuthasee (2000) note, “words that are familiar to the reader are likely to be interpreted quickly and correctly, requiring less cognitive energy for phonological analysis” (p. 6). Using this coding scheme, future research can investigate these details to further inform how students interpret items.

Second, future research needs to ask more explicit and detailed questions about the use of vignette illustrations in students’ experiences comprehending items. In this study, I only asked students if they saw the illustration and how they used it. Future investigations should explicitly ask questions about the illustration in relation to students’ understanding of the items. These questions should include students’ use of the illustration to understand specific words as well as
the entire stem. This research would then inform development of the illustration, focusing on important aspects to include or exclude from illustrations.

Finally, this study focused on multiple-choice science items. It is just the tip of the iceberg. Future studies utilizing this framework should be used to investigate different content areas (e.g., mathematics) as well as different item formats (e.g., constructed response). Detailed investigations of the issues presented in this dissertation across content areas and item formats would be invaluable to improving the assessment of linguistically diverse students. Continued use of the verbal protocol as a tool to investigate these issues is necessary to build a large base of data to inform the field about how emerging bilinguals interact with test items.

**Conclusion**

Accurate assessment is crucial for understanding what linguistically diverse students know. It should drive instruction and monitor the quality of education for emerging bilinguals. Currently, inaccurate measures are being used to make judgments about students, teachers, and schools. This study showed that there is great heterogeneity in the ways that emerging bilinguals make sense of and solve test items. This heterogeneity differed from the ways that mainstream monolinguals interacted with test items. It is clear that emerging bilinguals’ performance on multiple-choice science items is intricately intertwined with language. For this reason, test developers need to (a) consider linguistically diverse students from the beginning phases of test development, which requires the incorporation of theories of bilingualism and second language acquisition, (b) pilot test items with numerous emerging bilinguals to investigate how these items function across students, and (c) conduct analyses at a fine grain level to gain accurate insight to unmask differences across student groups.
References


Los Angeles: Dissemination and Assessment Center, California State University, 1-49.


*Educational Psychologist, 28*(1), 43-52.


Norwood, NJ: Ablex.


Educational testing service, Princeton, N. J.


*Linguistic and cultural influences on learning mathematics*, 201-220.


APPENDIX A

English verbal protocol

Prior to the interview:
- Check the recording device to make sure it works properly
- Set up the interview area: position chairs, set up papers
- Complete the Pre-Interview Form(s)
- Collect parent permission slip(s) from the teacher or student(s)

Interviewer instructions: If the student is a native Spanish speaker, incorporate Spanish into some of your introduction to let the student know you understand and speak Spanish.

- Give your name
- Ask student his/her name
- Thank the student for helping you with the interview.
  - Example script: Hi, my name is ___. I’m from the University of Colorado. I’m investigating how we can improve tests so that all students can better understand them. What is your name? Nice to meet you. Thank you for helping me with this interview.
- Explain the purpose of the interview and ensure the student this is NOT for a grade, create a relaxed environment.
  - Example script:
    - In a minute I will show you some questions. I want you to think out loud and tell me everything that you’re thinking about when answering the question. Do you know it means to think out loud? What I mean by think aloud is that I want you to say your thoughts out loud from the moment you read the question until you say the final answer. Try to be very specific and tell me as much as possible about what you’re thinking when you are figuring out the answer to the problem. Feel free to talk about anything that comes to mind, a feeling, a thought, anything you want. Pretend you are alone, talking to yourself. This is not for a grade and there is no right or wrong answer, I just want to know what you think when you are answering the question. I didn’t make the questions so you will not offend me with anything you say. Please feel free to say anything. If you are silent for a long time, I will remind you to think aloud. After you are finished, I will ask you to recall the steps you took to answer the question.
- Ask the student if he/she has any questions
- Ask the student if she understands

Warm-up Item #1 (Mathematics item-not a science item)
**Warm up: Concurrent Interview**

- Hand the student the warm-up item
- Ask him/her to please read the question out loud
- Ask him or her to think out loud
- Provide nonverbal communication
  - Example script: Now I’m going to hand you a question. I want you to read the question out loud and tell me everything you are thinking when you answer the question. I want to hear all your thoughts about what you are doing, what you are looking at, what you are thinking. Tell me as much as possible.

- Thank student for completing the item

**Warm-up #1: Retrospective interview**

- *Can you recall the steps you took to answer the question?*
  - Example script:
    - Can you tell me the steps that you did to answer the question? Tell me only the actual recalled thoughts you are confident that you had when you answered the question. Begin with the first thing you thought and did. Then tell me the next thought, and the next. Don’t worry about sounding repetitive, I want to know as many specific thoughts as you can.

**Warm-up Item # 2**

**Warm up: Concurrent Interview**

- Hand the student the warm-up item
- Ask him/her to please read the question out loud
- Ask him or her to think out loud
- Provide nonverbal communication
  - Example script: Now I’m going to hand you a question. I want you to read the question out loud and tell me everything you are thinking when you answer the question. I want to hear all your thoughts about what you are doing, what you are looking at, what you are thinking. Tell me as much as possible.

- Thank student for completing the item
Warm-up #2: Retrospective interview

- Can you recall the steps you took to answer the question?
  - Example script: Now I am going to ask you some questions about how you answered the question. I want you to tell me everything you were thinking and doing when you were answering the item. Can you tell me the steps that you did to answer the question? What was the first thing you did? Then what did you?

Warm-up #1: Closing questions:

- Did any part of the question help you solve the problem?
- What was the hardest part of the question?
- What was the question about? What was it asking you to do?
- Was there anything confusing about the question? What was it?

Warm-up #2: Closing questions:

- Did any part of the question help you solve the problem?
- What was the hardest part of the question?
- What was the question about? What was it asking you to do?
- Was there anything confusing about the question? What was it?

Item 1

Item 1 (Item with illustration): Concurrent Interview

- Hand item 1 to the student(s)
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem (ask pairs to work together)
- Provide nonverbal communication
- Thank student for completing the item

Item 1: Retrospective Interview: (For interviews with pairs, make sure BOTH students answer the question.)

- Can you recall the steps you took to answer the question?
Example script: Now I am going to ask you some questions about how you answered the question. I want you to tell me everything you were thinking and doing when you were answering the item. Can you tell me the steps that you did to answer the question? What was the first thing you did? Then what did you?

**Item 2**

**Item 2 (Item without illustration): Concurrent Interview**

- Hand item 2 to the student(s)
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem
- Provide nonverbal communication
- Thank student for completing the item

**Item 2: Retrospective Interview** (For interview with pairs: Make sure BOTH students answer the questions).

- *Can you recall the steps you took to answer the question?*
  - Example script: Now I am going to ask you some questions about how you answered the question. I want you to tell me everything you were thinking and doing when you were answering the item. Can you tell me the steps that you did to answer the question? What was the first thing you did? Then what did you?

**Item 3**

**Item 3 (Item with illustration): Concurrent Interview**

- Hand item 3 to the student
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem
- Provide nonverbal communication
- Thank student for completing the item

**Item 3: Retrospective Interview** (For interview with pairs: Make sure BOTH students answer the questions).

*Can you recall the steps you took to solve the problem?*
Item 4

Item 4 (Item without illustration): Concurrent Interview

- Hand item 4 to the student(s)
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem
- Provide nonverbal communication
- Thank student for completing the item

Item 4: Retrospective Interview (For interview with pairs: Make sure BOTH students answer the questions).

Can you please recall the steps you took to answer the question?

Item 1: Closing questions: (For interview with pairs: Make sure BOTH students answer the questions).
- Did any part of the question help you solve the problem?
- What was the hardest part of the question?
- If the student does not mention the illustration at all ask him/her:
  - Did you see the illustration on the page?
  - Did you use the illustration to help you solve the problem? Why or why not?
  - What did you think about when you saw the illustration?
- What was the question about? What was it asking you to do?

Was there anything confusing about the question? What was it?

Item 2: Closing questions: (For interview with pairs: Make sure BOTH students answer the questions).
- Did any part of the question help you solve the problem?
- What was the hardest part of the question?
- What was the question about? What was it asking you to do?

Was there anything confusing about the question? What was it?
Item 3: Closing questions:

- Did any part of the question help you solve the problem?

- What was the hardest part of the question?

- If the student does not mention the illustration at all ask him/her:
  - Did you see the illustration on the page?
  - Did you use the illustration to help you solve the problem? Why or why not?
  - What did you think about when you saw the illustration?

- What was the question about? What was it asking you to do?

Was there anything confusing about the question? What was it?

Item 4: Closing questions: (For interview with pairs: Make sure BOTH students answer the questions).

Did any part of the question help you solve the problem?

- What was the hardest part of the question?

- What was the question about? What was it asking you to do?

Was there anything confusing about the question? What was it?

End of interview questions (For interview with pairs: Make sure BOTH students answer the questions).

- How do you feel about answering these questions?

- How was your experience trying to answer these questions? Does anything stick out?

- Thanks the student for participating and ask if he/she has any questions.
APPENDIX B

Spanish verbal protocol

Prior to the interview:
- Check the recording device to make sure it works properly
- Set up the interview area: position chairs, set up papers
- Complete the Pre-Interview Form(s)
- Collect parent permission slip(s) from the teacher or student(s)

Introduction:

Interviewer instructions: If the student is a native Spanish speaker, incorporate Spanish into some of your introduction to let the student know you understand and speak Spanish.

- Give your name
- Ask student his/her name
- Thank the student for helping you with the interview.
  - Example script: Hola, me llamo ____, trabajo en la Universidad de Colorado. Estudio como se puede mejorar los exámenes para que todos los estudiantes puedan entenderlos. ¿Cómo te llamas? Mucho gusto. Gracias por ayudarme con la entrevista.
- Explain the purpose of the interview and ensure the student this is NOT for a grade, create a relaxed environment.
  - Example script:
    - Pronto te mostraré unas preguntas. Quiero que pienses en voz alta y que me digas todo lo que estás pensando. ¿Sabes que significa “pensar en voz alta?” Pensar en voz alta significa que dices tus pensamientos en voz alta desde leer la pregunta hasta que digas la respuesta. Intenta decir cosas específicas y decírmelo más que sea posible. Puedes decir cualquier cosa que se te ocurra. Por ejemplo, una emoción, un pensamiento, todo lo que quieres. Puedes actuar que estás solo(a) hablando con tu mismo. La entrevista no es para una nota, no hay respuestas correctas ni incorrectas. Solo quiero saber lo que piensas mientras respondes a la pregunta. No créas las preguntas así que no me vas a ofender con lo que dices. Si no dices nada te voy pedir que pienses en voz alta. Después de que terminas, te voy a pedir contar los pasos que hiciste para responder a la pregunta.
- Ask the student if he/she has any questions. ¿Tienes alguna pregunta?
- Ask the student if she understands ¿Entendiste lo que te he explicado?

Warm-up Item # 1 (Mathematics item-not a science item)
Warm up: Concurrent Interview

- Hand the student the warm-up item
- Ask him/her to please read the question out loud
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.
- Provide nonverbal communication. If the student is silent, ask him/her to think aloud. Habla.
- Thank student for completing the item. Gracias por hacer la pregunta.

Warm-up #1: Retrospective interview

- ¿Me cuentas los pasos que hiciste para responder a la pregunta?
  - Example script:
    - Me cuentas los pasos que hiciste para responder a la pregunta? Empieza con lo que hiciste primero. Y después, dime el próximo pensamiento, y después el próximo. No importa si te repites, quiero saber los detalles de tus pensamientos.

Warm-up Item # 2

Warm up: Concurrent Interview

- Hand the student the warm-up item
- Ask him/her to please read the question out loud
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.
- Ask him or her to think out loud Habla.
- Provide nonverbal communication
- Thank student for completing the item. Gracias por hacer la pregunta.

Warm-up #2: Retrospective interview

- Can you recall the steps you took to answer the question?
  - Example script: Cuéntame los pasos que hiciste para responder a la pregunta. Empieza con lo que hiciste primero. Y después, dime el próximo
pensamiento, y después el próximo. No importa si te repites, quiero saber los detalles de tus pensamientos.

**Warm-up #1: Closing questions:**

- ¿Algo te ayudó en responder a la pregunta?
- ¿Qué fue la parte más difícil de la pregunta?
- ¿De qué se trata la pregunta? ¿Qué te pidió hacer?
- ¿Te confundió alguna parte de la pregunta? ¿Qué fue?

**Warm-up #2: Closing questions:**

- ¿Algo te ayudó en responder a la pregunta?
- ¿Qué fue la parte más difícil de la pregunta?
- ¿De qué se trata la pregunta? ¿Qué te pidió hacer?
- ¿Te confundió alguna parte de la pregunta? ¿Qué fue?

**Item 1**

**Item 1 (Item with illustration): Concurrent Interview**

- Hand item 1 to the student(s)
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem (ask pairs to work together)
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, qué piensas. Dime lo más que sea posible.
- Provide nonverbal communication. If the student is silent, ask him/her to keep talking. *Habla.*
- Thank student for completing the item

**Item 1: Retrospective Interview:** (For interviews with pairs, make sure BOTH students answer the question.)

- Cuéntame los pasos que hiciste para responder a la pregunta.
Example script: Cuéntame los pasos que hiciste para responder a la pregunta? Empieza con lo que hiciste primero. Y después, dime el próximo pensamiento, y después el próximo. No importa si te repites, quiero saber los detalles de tus pensamientos.

**Item 2**

**Item 2 (Item without illustration): Concurrent Interview**

- Hand item 2 to the student(s)
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.
- Provide nonverbal communication. If the student is silent, ask him/her to keep talking. *Habla.*
- Thank student for completing the item

**Item 2: Retrospective Interview** (For interview with pairs: Make sure BOTH students answer the questions).

- Cuéntame los pasos que hiciste para responder a la pregunta.
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.

**Item 3**

**Item 3 (Item with illustration): Concurrent Interview**

- Hand item 3 to the student
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.
- Provide nonverbal communication. If the student is silent, ask him/her to keep talking. *Habla.*
Thank student for completing the item

**Item 3: Retrospective Interview** (For interview with pairs: Make sure BOTH students answer the questions).

- *Cuéntame los pasos que hiciste para responder a la pregunta.*
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.

**Item 4**

**Item 4 (Item without illustration): Concurrent Interview**

- Hand item 4 to the student(s)
- Ask him/her to read the item aloud
- Ask him/her to think out loud while he/she solves the problem
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.

  Provide nonverbal communication. If the student is silent, ask him/her to keep talking. *Habla.*

- Thank student for completing the item

**Item 4: Retrospective Interview** (For interview with pairs: Make sure BOTH students answer the questions).

- *Cuéntame los pasos que hiciste para responder a la pregunta.*
  - Example script: Ahora te doy una pregunta. Quiero que leas la pregunta en voz alta y digas todo lo que piensas mientras respondes a la pregunta. Quiero escuchar todo lo que haces, donde miras, que piensas. Dime lo más que sea posible.

**Item 1: Closing questions:** (For interview with pairs: Make sure BOTH students answer the questions).

- ¿Algo te ayudó en responder a la pregunta?
- ¿Qué fue la parte más difícil de la pregunta?

If the student does not mention the illustration at all ask him/her:
  - ¿Viste la imagen?
  - ¿Qué pensaste cuando viste la imagen?
¿La imagen te ayudó responder a la pregunta? Por qué?

- ¿De qué se trata la pregunta? ¿Qué te pidió hacer?
- ¿Te confundió alguna parte de la pregunta? ¿Qué fue?

**Item 2: Closing questions:** (For interview with pairs: Make sure BOTH students answer the questions).

- ¿Algo te ayudó en responder a la pregunta?
- ¿Qué fue la parte más difícil de la pregunta?
- ¿De qué se trata la pregunta? ¿Qué te pidió hacer?

¿Te confundió alguna parte de la pregunta? ¿Qué fue?
Item 3: Closing questions:

- ¿Algo te ayudó en responder a la pregunta?
- ¿Qué fue la parte más difícil de la pregunta?
- If the student does not mention the illustration at all ask him/her:
  - ¿Viste la imagen?
  - ¿Qué pensaste cuando viste la imagen?
  - ¿La imagen te ayudó responder a la pregunta? ¿Por qué?
- ¿De qué se trata la pregunta? ¿Qué te pidió hacer?
- ¿Te confundió alguna parte de la pregunta? ¿Qué fue?

Item 4: Closing questions: (For interview with pairs: Make sure BOTH students answer the questions).

- ¿Algo te ayudó en responder a la pregunta?
- ¿Qué fue la parte más difícil de la pregunta?
- ¿De qué se trata la pregunta? ¿Qué te pidió hacer?
- ¿Te confundió alguna parte de la pregunta? ¿Qué fue?

End of interview questions (For interview with pairs: Make sure BOTH students answer the questions).

- ¿Cómo sentías cuando respondías a las preguntas?
- ¿Cómo lo pasaste intentando responder a las preguntas? ¿Algo sobresale?
- Muchas gracias por participar en la entrevista. ¿Tienes alguna pregunta?
APPENDIX C

Classifying science terms

Please read the following 10 items and identify in each science–related terms in both the stem and the options. Science-related terms refer to objects, processes, concepts, theories, or symbols used to discuss scientific content. They also are terms that are not specific to science but are likely to be part of the language used in science because of their precision.

Identify science-related terms by circling and coding with a number each term according the categories given below.

**Coding Categories**

<table>
<thead>
<tr>
<th>Code</th>
<th>Type of term</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Terms specific to the domain of science; only used when discussing science</td>
<td>electron, refraction</td>
</tr>
<tr>
<td>2</td>
<td>Terms used in science but is also common to everyday conversations</td>
<td>energy, weight</td>
</tr>
<tr>
<td>3</td>
<td>Terms with only one meaning</td>
<td>gain, optimum</td>
</tr>
<tr>
<td>4</td>
<td>Terms with more than one meaning</td>
<td>negative, light, mass</td>
</tr>
<tr>
<td>5</td>
<td>Terms that appear frequently in the scientific discourse</td>
<td>likely, in contrast</td>
</tr>
</tbody>
</table>

Please note that a science term can consist of more than one word. If a term includes more than one word, please identify individual science words as well as the entire science term. For example, the term, *line of symmetry* is identified as a mathematical term. In addition, *symmetry* is identified as a mathematical word. Therefore, both *symmetry* and *line of symmetry* should be circled and coded:

```
1, 3
line of symmetry
```

**Example** (Rachel’s non-expert coding): Please circle and code as shown in the example below.

**Original item:**

For a special dinner, Catherine’s mom lit some candles in the living room for decoration. What two forms of energy does the fire from a burning candle release?

A. Light and heat  
B. Sound and chemical  
C. Magnetic and nuclear  
D. Electrical and mechanical
For a special dinner, Catherine’s mom lit some candles in the living room for decoration. What two forms of energy does the fire from a burning candle release?

A. Light and heat

B. Sound and chemical

C. Magnetic and nuclear

D. Electrical and mechanical