The Development and Validation of the Middle School-Life Science Concept Inventory (MS-LSCI) Using Rasch Analysis

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Andria Stammen, M.A., M.Ed.
Graduate Program in Education Teaching & Learning

The Ohio State University
2018

Dissertation Committee:
Karen E. Irving, Ph.D., Chair
Kathy Malone, Ph.D.
William Boone, Ph.D.
Zakee Sabree, Ph.D.
Abstract

The aim of this research is to develop a measurement instrument that is valid and reliable, called the Middle School-Life Science Concept Inventory (MS-LSCI), for the purpose of measuring the life science conceptual understanding of middle school-level students. Although there are several existing concept inventories related to biology concepts (i.e. Secondary-Biology Concept Inventory (S-BCI) and Biology Concept Inventory (BCI)), there is no fully developed concept inventory available that collectively measures the major life science concepts covered in middle school classrooms (Klymkowsky, Underwood, & Garvin-Doxas, 2010; Stammen, Lan, Schuchaerdt, Malone, Ding, Sabree, & Boone, 2016). Study one focuses on how data from a multi-panel expert review and student interviews were used in the Middle School-Life Science’s (MS-LSCI) item content qualitative validation and iterative refinement process. Of the 50 questions reviewed by the expert panels, 12 items were identified as having content validity concerns. Generally, these content validity concerns fell within two categories: (i) imprecise phrasing and (ii) age inappropriateness. During the student interviews, a total of 26 items were identified as displaying content validity issues. These 26 items fell into one of three categories: (i) imprecise phrasing, (ii) contextual ambiguity, and (iii) formatting/diagrammatic complexity. Using the data from the multi-panel expert review and student interviews, the items with content validity concerns were
refined and modified before the items were field tested. Study two describes the MS-LSCI’s quantitative validation and item selection process. Specifically, this study focuses on the psychometric functioning of the 60 field-tested MS-LSCI items using Rasch analysis. The results of this development, refinement, and evaluation process suggest that the 25-item MS-LSCI is a valid instrument in that the items appear to be unidimensional, item and person measures display little misfit, and the reliability values suggest replicability within the targeted sample. Based on these results, the MS-LSCI has the potential to help fill the gap in the assessment tools available to measure middle school student life science concepts in conjunction with alternative conceptions. Study three describes the difference among 6th, 7th, and 8th graders’ performance on the MS-LSCI’s 60 field tested items. The results suggest that MS-LSCI performance was significantly lower for 6th graders relative to 8th graders and that six items can be linked to this difference in performance. These six items fell within the MS-LSCI’s core concepts of evolution and diversity, population interaction, and growth and reproduction. More importantly, these six items shed light on the contextual differences between 6th and 8th grade performance. The contextualized item differences discovered in this study may be attributed to middle school life science curriculum sequencing and the perseverance of specific alternative conceptions.
Dedication

Traveler there is no path.
The path forms itself as you walk it.
-Antonio Machado

This work is dedicated to my best friend Josh, dad Charlie, mom Denise, and sisters Vanessa and Allison for walking with me to form this path. With unconditional love and sacrifices, you have granted me this epic journey that would change everything.
Acknowledgments

The path to completing a doctorate is one not walked alone. First and foremost, I would like to extend my gratitude to my co-advisors, Dr. Kathy Malone and Dr. Karen Irving. To Dr. Malone, thank you for being the visionary who helped me find my calling in science education research. Your mentorship has afforded me so many learning opportunities. To Dr. Irving, thank you for your support and guidance while learning how to navigate the research enterprise. Your commitment to preservice teacher education is inspiring. To my committee member Dr. Bill Boone, your enthusiasm for Rasch analysis is contagious, and I am grateful for our conversations about what this psychometric model has to offer science education research. To my committee member Dr. Zakee Sabree, thank you for your willingness to share your content knowledge expertise, insights on the research community, and your words of encouragement.

I cannot go without thanking the faculty and staff I had the honor of learning from while completing my graduate and doctoral coursework. To Gail Hoskins who made a lasting impact on my educational and professional experiences and goals, thank you for believing in me and allowing me to ‘get into your head’ to learn. To Tami Augustine and program managers, I have learned so much from you, and your continued guidance and support should not go without recognition.

Thank you to my classmates, whom I have had the great privilege of working with over the past four year. To Leah Frazee and Katie Miller, who were always there to provide humor and support in times of doubt.

There are so many teachers and students to thank, including those who served as expert panelists, participated in interviews, and administered the MS-LSCI. To BW, without you, I would not have met ‘the’ timeline. Forever shall I be indebted to all of you.

To my family, who has been and will always be the pillars of strength, motivation, and support in my life. Mere words cannot express the admiration I have for parents Denise Stammen and Charlie Stammen who have instilled in me the value of hard work and perseverance, along with the belief that actions often speak louder than words. Thank you for always believing that I can do anything I set my heart and mind to. To Allison Stammen and Vanessa Stammen, you have always been there through thick and thin, and I am beyond lucky to have you as sisters, now and forever. To Josh Roth, who has been my best friend and rock for over 14 years, I promise you will have me back from this day forward. Thank you for your patience, support, and humor throughout this journey. To Frankie, Max, and Wiley, thank you for your limitless snuggles and wild antics that continue to bring endless laughter and joy. I love you all to the ends of the earth and back again.
Vita

2008..............................B.A. Biological Sciences, Wright State University

2009..............................M.Ed. Adolescent to Young Adult Life Science, Wright State University

2017..............................M.A. Educational Studies, The Ohio State University

2014 to 2016 .........................Graduate Teaching and Research Associate, Department of Teaching and Learning, The Ohio State University

2016 to present .....................STEM Education Program Manager, Department of Teaching and Learning, The Ohio State University
Publications


Fields of Study

Major Field: Education Teaching & Learning

Science, technology, Engineering, and Mathematics (STEM) Education
Table of Contents

Abstract ....................................................................................................................................... ii

Dedication ................................................................................................................................... iv

Acknowledgments ...................................................................................................................... v

Vita .............................................................................................................................................. vi

Publications ............................................................................................................................... vii

Fields of Study .......................................................................................................................... vii

Table of Contents ..................................................................................................................... viii

List of Tables ............................................................................................................................ xvii

List of Figures ............................................................................................................................. xx

Chapter 1: Introduction and Statement of the Problem ............................................................... 1

Overview ..................................................................................................................................... 1

Background and Conceptual Framework .................................................................................... 1

Concept Inventories ................................................................................................................... 1

Alternative Conceptions ........................................................................................................... 2

Common Alternative Conceptions in Biology ........................................................................... 3
Evolution and Natural Selection ................................................................. 4
Cells ................................................................................................................. 9
Human Body Systems ..................................................................................... 10
Interdependence in Ecosystems ..................................................................... 14
Matter and Energy .......................................................................................... 19
Reproduction, Genes, and Heredity ............................................................... 22
Curriculum and Instructional Strategies Targeting Alternative Conceptions ....... 25
Assessments Targeting Alternative Conceptions .............................................. 27
Three Key Pedagogical Characteristics of Assessment ...................................... 28
Overview of Life Science Concept Inventories ................................................. 30
Tertiary-Level (College-Level) Concept Inventories ......................................... 32
Experimental Design ....................................................................................... 34
Evolution and Natural Selection ..................................................................... 34
Energy and Matter .......................................................................................... 36
Inheritance ...................................................................................................... 36
Molecular Biology ............................................................................................ 37
Comprehensive Biology Assessments .............................................................. 38
Theoretical Framework ................................................................. 61

Qualitative Design and Validation Process ..................................... 61

Research Goals ............................................................................. 63

Methods ....................................................................................... 63

Summary of the design and validation process for the MS-LSCI .......... 64

Test Philosophy: Instrument Structure and Constructs .................... 64

Item Development ........................................................................ 67

Item Review .................................................................................. 69

Expert panels .............................................................................. 69

Student Interviews. ...................................................................... 69

Analysis ....................................................................................... 70

Expert Panel Review ................................................................. 71

Exemplar I: Age Inappropriateness ............................................. 72

Exemplar II: Imprecise Phrasing .................................................... 75

Student Interview Review ............................................................ 78

Exemplar III: Imprecise Phrasing ................................................. 79

Exemplar IV: Contextual Ambiguity ............................................. 82
Exemplar V: Formatting/Diagrammatic Complexity ........................................... 87

Results ................................................................................................................. 93

Discussion and Conclusion .................................................................................. 95

Future Steps .......................................................................................................... 96

Implications ........................................................................................................... 97

Limitations ............................................................................................................ 97

References ............................................................................................................ 99

Chapter 3: Study Two ........................................................................................... 103

The Development and Validation of the Middle School-Life Science Concept

Inventory (MS-LSCI): A Quantitative Study Using Rasch Modeling ................... 103

Abstract .............................................................................................................. 103

Introduction ......................................................................................................... 103

Theoretical Framework ......................................................................................... 105

Quantitative Design and Validation Process ....................................................... 105

Research Goal ...................................................................................................... 107

Methods ............................................................................................................... 107

Test Philosophy Framework ................................................................................. 107
Item Development ........................................................................................................ 109

Item Review: Expert Panels and Student Interviews............................................. 110

Field-testing Data Collection..................................................................................... 114

Rasch Analysis ............................................................................................................. 115

Results .......................................................................................................................... 116

Local Independence and Unidimensionality ............................................................. 116

Item Difficulty and Person Ability ........................................................................ 118

Item Selection for the Final 25-Item MS-LSCI...................................................... 121

Reliability .................................................................................................................... 124

Discussion and Conclusion ....................................................................................... 125

Limitations .................................................................................................................. 126

Implications ................................................................................................................ 127

Future Research ........................................................................................................ 128

References ............................................................................................................... 129

Chapter 4: Study Three .............................................................................................. 133

A Contextualized Item Comparison Study of Middle Schooler Performance on the
Middle School-Life Science Concept Inventory (MS-LSCI) .................................. 133
Discussion .................................................................................................................. 152

Connecting Contextualized Item Difference and Response Patterns to Middle School Life Science Curriculum and Alternative Conceptions ........................................... 153

Evolution and diversity ............................................................................................... 154

Population interactions .............................................................................................. 155

Growth and reproduction ........................................................................................... 156

Implications ................................................................................................................ 157

Future Research ......................................................................................................... 158

Limitations ................................................................................................................ 158

References .................................................................................................................. 160

Chapter 5: Discussion of Three Studies ...................................................................... 162

Overview of Studies .................................................................................................... 162

Implications ................................................................................................................ 163

Pre-assessment .......................................................................................................... 163

Pre/post assessment .................................................................................................. 164

Post-assessment ......................................................................................................... 165

Additional Utility ....................................................................................................... 166
Limitations ................................................................................................................. 167
Future Research ......................................................................................................... 168
References .................................................................................................................. 171
Complete References ................................................................................................ 173
List of Tables

Table 1. Categories of alternative conceptions (Committee on Undergraduate Science Education (U.S.), 1997). ........................................................................................................... 3

Table 2. Key ideas about evolution and natural selection (AAAS Project 2061, n.d.; NGSS Lead States, 2013). ........................................................................................................... 5

Table 3. Seven common middle school alternative concepts related to evolution and natural selection (AAAS, n.d.). ........................................................................................................... 7

Table 4. Key ideas about cells (AAAS Project 2061, n.d.; NGSS Lead States, 2013). .... 9

Table 5. Seven common middle school alternative concepts related to cells (AAAS, n.d.). ........................................................................................................... 10

Table 6. Key ideas about human body systems (AAAS Project 2061, n.d.; NGSS Lead States, 2013). ........................................................................................................... 12

Table 7. Seven common middle school alternative concepts related to human body systems (AAAS, n.d.). ........................................................................................................... 13

Table 8. Key ideas about interdependence in ecosystems (AAAS Project 2061, n.d.; NGSS Lead States, 2013). ........................................................................................................... 15

Table 9. Seven common middle school alternative concepts related to the interdependence in ecosystems (AAAS, n.d.). ......................................................................................... 17
Table 10. Key ideas about matter and energy in living systems (AAAS Project 2061, n.d.; NGSS Lead States, 2013).

Table 11. Seven common middle school alternative concepts related to matter and energy in living systems (AAAS, n.d.).

Table 12. Key ideas about reproduction, genes, and heredity (AAAS Project 2061, n.d.; NGSS Lead States, 2013).

Table 13. Seven common middle school alternative concepts related to reproduction, genes, and heredity (AAAS, n.d.).

Table 14. Summary of Life Science Concept Inventories.

Table 15. Summary of Tertiary-Level Life Science Concept Inventories.

Table 16. Summary of Secondary-Level Life Science Concept Inventories.

Table 17. Essential questions associated with the MS-LSCI core concepts (NGSS Lead States, 2013; Stammen, et al, 2016).

Table 18. Number of items per MS-LSCI core concepts.

Table 19. Distribution of items with content validity concerns by student interview group.

Table 20. Number of items per MS-LSCI core concepts used during field testing.

Table 21. Essential questions associated with the MS-LSCI core concepts (NGSS Lead States, 2013; Stammen, et al., 2016).

Table 22. MS-LSCI item description.
Table 23. MS-LSCI logit mean performance scores per grade level......................... 139

Table 24. Grade level performance means to logit mean conversion.......................... 141

Table 25. Crosswalk between MS-LSCI items, NGSS Disciplinary Core Ideas, and MS-LSCI core concepts. ........................................................................................................... 143

Table 26. Item 3 response pattern.............................................................................. 145

Table 27. Item 4 response pattern.............................................................................. 147

Table 28. Item 15 response pattern............................................................................ 148

Table 29. Item 19 response pattern............................................................................ 149

Table 30. Item 20 response pattern............................................................................ 150

Table 31. Item 25 response pattern............................................................................ 152
List of Figures

Figure 1. Overview of the theoretical framework of scientific constructivism........... 26

Figure 2. Quality formative assessment practices (Black & William, 2010)................ 29

Figure 3. Summary of the design and validation process for a concept inventory. ........ 62

Figure 4. Summary of the design and validation process for the MS-LSCI (Note: * represents when item revision occurred). ........................................................................................................ 64

Figure 5. MS-LSCI sample item highlighting the alternative conception distractor choices [item adapted from (Stamp, 2015)]. ................................................................ 68

Figure 6. Summary of the MS-LSCI’s item review and revision (arrows represent item revision). .................................................................................................................................................. 71

Figure 7. Original item displaying age inappropriateness........................................ 73

Figure 8. Edited item displaying age inappropriateness (modifications are highlighted). 74

Figure 9. Original item displaying imprecise phrasing.............................................. 75

Figure 10. Edited item displaying imprecise phrasing (modifications are highlighted). 76

Figure 11. Original and adapted population interaction items (Stamp, 2015). .......... 77

Figure 12. Original and adapted energy and matter items.................................... 77

Figure 13. Original item displaying imprecise phrasing........................................... 80

Figure 14. Post-undergraduate student interviews edited item displaying imprecise phrasing.................................................................................................................................................. 81
Figure 15. Post-middle school interviews edited item displaying imprecise phrasing (modifications are highlighted). ................................................................. 81

Figure 16. Original item displaying contextual ambiguity ........................................ 83

Figure 17. Post-undergraduate student interviews edited item displaying contextual ambiguity (modifications are highlighted) .............................................................. 84

Figure 18. Post-middle school interviews edited item displaying contextual ambiguity (modifications are highlighted) ................................................................. 86

Figure 19. Original item displaying formatting/diagrammatic complexity ................ 88

Figure 20. Post-undergraduate student interviews edited item displaying formatting/diagrammatic complexity (modifications are highlighted) .............................................................. 90

Figure 21. Post-middle student interviews edited item displaying formatting/diagrammatic complexity (modifications are highlighted) .............................................................. 92

Figure 22. Original and adapted inheritance items .................................................. 93

Figure 23. Summary of the MS-LSCI’s item review and revision (arrows represent item revision). .................................................................................................................. 94

Figure 24. Summary of the design and validation process for a concept inventory. ..... 106

Figure 25. MS-LSCI sample field-tested item highlighting the alternative conception distractor choices. ................................................................................................................................. 110

Figure 26. Example of a MS-LSCI field-testing additional item. ............................... 113
Figure 27. Wright map for 60 item MS-LSCI. Wright map for 60 item MS-LSCI. The left side of the map shows the distribution of student ability and the right side shows the distribution of item difficulty. Items are labeled A1-A60. The labels M, S, and T represent the mean value, one standard deviation, and 2 standard deviations for each distribution, respectively. Each ‘#’ represents 6 students and each ‘.’ represents between 1 and 5 students.

Figure 28. Wright map for the 25 item MS-LSCI. The left side of the map shows the distribution of student ability and the right side shows the distribution of item difficulty. The labels M, S, and T represent the mean value, one standard deviation, and 2 standard deviations for each distribution, respectively. Each ‘#’ represents 8 students and each ‘.’ represents between 1 and 7 students.

Figure 29. Wright map for 60 item MS-LSCI with the 6th, 7th, and 8th grade level logits means are plotted at -.86, -.76, and -.69 logits, respectively. The left side of the map shows the distribution of student ability and the right side shows the distribution of item difficulty. Items are labeled A1-A60. The labels M, S, and T represent the mean value, one standard deviation, and 2 standard deviations for each distribution, respectively. Each ‘#’ represents 5 students and each ‘.’ represents between 1 and 4 students.

Figure 30. MS-LSCI item 3

Figure 31. MS-LSCI item 4

Figure 32. MS-LSCI item 15
Figure 33. MS-LSCI item 19. .......................................................... 149

Figure 34. MS-LSCI item 20. .......................................................... 150

Figure 35. MS-LSCI item 25. .......................................................... 152
Chapter 1: Introduction and Statement of the Problem

Overview

The aim of this research is to develop a measurement instrument that is shown to be valid and reliable, called the Middle School-Life Science Concept Inventory (MS-LSCI), for measuring life science conceptual understanding of middle school-level students. Although there are several existing concept inventories related to biology concepts (i.e. Secondary-Biology Concept Inventory (S-BCI) and Biology Concept Inventory (BCI)), there is no fully developed concept inventory available that collectively measures the major life science concepts covered in middle school classrooms (Klymkowsky, Underwood, & Garvin-Doxas, 2010; Stammen, Lan, Schuchaerd, Malone, Ding, Sabree, & Boone, 2016).

Background and Conceptual Framework

*Concept Inventories*

Concept inventories are empirically developed instruments designed to measure conceptual understanding of concepts for which students share “alternative conceptions and faulty reasoning” (D'Avanzo, 2008, p. 1; Hestenes, Wells, Swackhamer, 1992). These assessments can be useful tools in measuring what students have learned in science and/or used to evaluate the implementation of newly developed science curriculum (Liu, 2010). In general, concept inventories display a selected response test format. Unlike
traditional selected response tests, a concept inventory’s item distractors (i) are grounded in extensive research, (ii) are phrased using student explanations, (iii) diagnose a specific level of student understanding, and (iv) aid in revealing students’ alternative conceptions (Garvin-Doxas, Klymkowski, & Elrod, 2007; D'Avanzo, 2008, p. 2; Tamir, 1971). Thus, unlike traditional selected response tests that only provide information on the quantity of students who incorrectly selected a response for each item, concept inventories can help provide researchers and teachers a conceptual snapshot of how each student is understanding concepts along with alternative conceptions students possess.

**Alternative Conceptions**

Alternative conceptions, referred to as misconceptions in prior literature, are pervasive and persistent misunderstandings that are not aligned with expert-like conceptual understanding in science (McDermott, 1991). That is, alternative conceptions are held by a large proportion of students and are highly resistant to instruction (Marbach-Ad, Briken, El-Sayed, Frauwirth, Fredericksen, Hutcheson, … Smith, 2009). Alternative conceptions can be classified into one of the five following categories: preconception, nonscientific belief, conceptual misunderstanding, vernacular misconception, and factual misconception (refer to Table 1 for additional details related these categories of alternative conceptions) (Committee on Undergraduate Science Education (U.S.), 1997). Science education literature offers a large body of research that describes students’ alternative conceptions in all science domains including those related
to life science topics. This research is a foundational component to the development of concept inventories as item distractors are often linked to alternative conceptions (Tamir, 1971; Treagust, 1986).

Table 1. Categories of alternative conceptions (Committee on Undergraduate Science Education (U.S.), 1997).

<table>
<thead>
<tr>
<th>Alternative conception category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preconception</td>
<td>Arise from popular conceptions rooted in everyday experiences</td>
</tr>
<tr>
<td>Nonscientific belief</td>
<td>Views learned by students from sources other than scientific education, such as religious or mythical teachings</td>
</tr>
<tr>
<td>Conceptual misunderstanding</td>
<td>Arise when students are taught scientific information in a way that does not provoke them to confront paradoxes and conflicts resulting from their own preconceived notions and nonscientific beliefs</td>
</tr>
<tr>
<td>Vernacular misconception</td>
<td>Arise from the use of words that mean one thing in everyday life and another in a scientific context (e.g., &quot;work&quot;)</td>
</tr>
<tr>
<td>Factual misconception</td>
<td>Falsities often learned at an early age and retained unchallenged into adulthood (e.g. lighting never strikes twice in the same place)</td>
</tr>
</tbody>
</table>

Common Alternative Conceptions in Biology

Because subsequent chapters will involve the development of a concept inventory in the context of middle school, this section will highlight major alternative conceptions related to life science in grades 6 through 8. The American Association for the Advancement of Science’s (AAAS) Project 2061, in part, offers a synopsis of many of
the identified alternative conceptions held by middle schoolers. According to the AAAS Project 2061, life science content can be divided into six topics: (i) evolution and natural selection; (ii) cells; (iii) human body systems; (iv) interdependence in ecosystems; (v) matter and energy in living systems; and (vi) reproduction, genes, and heredity (AAAS Project 2061, n.d.). According to AAAS, the “frequency of selecting a misconception was calculated by dividing the total number of times a misconception was chosen by the number of times it could have been chosen, averaged over the number of students answering the questions within this particular idea.” The most common alternative conceptions will be explored for each of the six life science concepts if over 15 percent of the students measured by AAAS held an alternative conceptions. In addition, alternative conceptions identified in science education literature will be described for each of the six life science topics listed above.

**Evolution and Natural Selection.** Evolution and natural selection can be divided into four key ideas (refer to Table 2) (AAAS Project 2061, n.d.). This table also relates these key ideas about evolution and natural selection to the Next Generation Science Standards, which are often used to guide secondary classroom curriculum including assessments (NGSS Lead States, 2013).
Table 2. Key ideas about evolution and natural selection (AAAS Project 2061, n.d.; NGSS Lead States, 2013).

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Description</th>
<th>NGSS Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>There are similarities and differences between organisms living today and those that lived in the past.</td>
<td>MS-LS4-1: Analyze and interpret data for patterns in the fossil record that document the existence, diversity, extinction, and change of life forms throughout the history of life on Earth under the assumption that natural laws operate today as in the past.</td>
</tr>
<tr>
<td>2</td>
<td>Environmental conditions have changed in the past and continue to change today.</td>
<td>MS-LS4-1: Analyze and interpret data for patterns in the fossil record that document the existence, diversity, extinction, and change of life forms throughout the history of life on Earth under the assumption that natural laws operate today as in the past.</td>
</tr>
<tr>
<td>3</td>
<td>When inherited traits are favorable to individual organisms, the proportion of individuals in a population that have those traits will tend to increase over successive generations.</td>
<td>MS-LS4-4: Construct an explanation based on evidence that describes how genetic variations of traits in a population increase some individuals’ probability of surviving and reproducing in a specific environment.</td>
</tr>
<tr>
<td>4</td>
<td>Similarities and differences in inherited characteristics of organisms alive today or in the past can be used to infer the relatedness of any two species, change in species over time, and lines of evolutionary descent.</td>
<td>MS-LS4-2: Apply scientific ideas to construct an explanation for the anatomical similarities and differences among modern organisms and between modern and fossil organisms to infer evolutionary relationships.</td>
</tr>
</tbody>
</table>

Each of the four key ideas about evolution and natural selection has between 4 and 11 corollary alternative conceptions (AAAS Project 2061, n.d). These alternative
conceptions were measured for prevalence at both the middle school level (grades 6 to 8) and high school level (grades 9 to 12). Table 3 describes the seven most common middle school alternative conceptions reported by AAAS (n.d.).
Table 3. Seven common middle school alternative concepts related to evolution and natural selection (AAAS, n.d.).

<table>
<thead>
<tr>
<th>Key idea</th>
<th>Alternative conception</th>
<th>Percent frequency of selecting alternative conception</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Species that have no apparent, obvious or superficial similarities have no similarities at all.</td>
<td>39%</td>
</tr>
<tr>
<td>3</td>
<td>Individual organisms can deliberately develop new heritable traits because they need them for survival (Bishop &amp; Anderson, 1990; Passmore &amp; Stewart, 2002).</td>
<td>34%</td>
</tr>
<tr>
<td>3</td>
<td>Sudden environmental change is required for evolution to occur (Nehm &amp; Reilly, 2007).</td>
<td>32%</td>
</tr>
<tr>
<td>4</td>
<td>Changes in a population occur through a gradual change in all members of a population, not from the survival of a few individuals that preferentially reproduce (Brumbly, 1979; Bishop &amp; Anderson, 1990; Anderson, Fisher, &amp; Norman, 2002).</td>
<td>26%</td>
</tr>
<tr>
<td>4</td>
<td>Species that are similar can share a common ancestor, but species that have no apparent, obvious, or superficial similarities cannot share a common ancestor.</td>
<td>47%</td>
</tr>
<tr>
<td>4</td>
<td>Plants and animals cannot share a common ancestor (Bizzo, 1994; Ha &amp; Cha, 2008).</td>
<td>40%</td>
</tr>
<tr>
<td>4</td>
<td>Humans do not share a common ancestor with other living organisms (Ha &amp; Cha, 2008; Stern &amp; Hagay, 2005).</td>
<td>35%</td>
</tr>
</tbody>
</table>

Within the topic of evolution and natural selection, science education literature identifies additional alternative conceptions related to Lamarckian principles, organisms consciously deciding to change, and what counts as evolutionary fitness. Several
evolution and natural selection alternative conceptions are linked to an organism’s conscious intention to change. Refer to the following examples: (i) natural selection involves organisms trying to adapt (Understanding Evolution, 2015); (ii) mutations occur to meet the needs of the population (Battisti, Hanegan, Sudweeks, & Cates, 2010); (iii) evolution results in progress, meaning that organisms are always getting better through evolution (Understanding Evolution, 2015); (iv) individual organisms or populations of organisms will develop the traits they need (Battisti et al., 2010); (v) changes in traits are need driven and mutations are intentional, and therefore randomness is not important (Anderson et al., 2002; Klymkowsky, Garvin-Doxas, & Zeilik, 2003); and (vi) adaptations are due to an overall purpose of conscious process (Soderberg & Price, 2003). Other ubiquitous evolution and natural selection alternative conceptions are related to Lamarckian principles. Refer to the following examples: (i) when parents develop stronger claws through repeated use in catching prey, their offspring can inherit the stronger claw trait (Battisti et al., 2010); and (ii) individually acquired adaptations can be passed onto offspring (inheritable phenotypic change) (Abraham, Perez, & Price, 2014; Soderberg & Price, 2003). Alterative conceptions related to fitness are common in science education literature. Refer to the following examples: (i) the fittest organisms in a population are those that are strongest, healthiest, fastest, and/or largest; (ii) natural selection is about survival of the very fittest individuals in population; and (iii) individual organisms can evolve during a signal lifespan (Understanding Evolution, 2015).
**Cells.** The life science topic of cells can be divided into four key ideas (refer to Table 4) (AAAS Project 2061, n.d.). This table also summarizes the connections between these key ideas to the Next Generation Science Standards (NGSS Lead States, 2013).

**Table 4. Key ideas about cells (AAAS Project 2061, n.d.; NGSS Lead States, 2013).**

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Description</th>
<th>NGSS Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All living things are composed of one or more cells.</td>
<td>MS-LS1-1: Conduct an investigation to provide evidence that living things are made of cells; either one cell or many different numbers and types of cells.</td>
</tr>
<tr>
<td>2</td>
<td>Although there are many different types of cells in terms of size, structure, and function, all cells have certain characteristics in common.</td>
<td>MS-LS1-2: Develop and use a model to describe the function of a cell as a whole system and ways parts of cells contribute to the function.</td>
</tr>
<tr>
<td>3</td>
<td>Cells in multicellular organisms repeatedly divide to make more cells for growth and repair.</td>
<td>MS-LS1-3: Use argument supported by evidence for how the body is a system of interacting subsystems composed of groups of cells.</td>
</tr>
<tr>
<td>4</td>
<td>Different body structures are made up of different types of cells.</td>
<td>MS-LS1-3: Use argument supported by evidence for how the body is a system of interacting subsystems composed of groups of cells.</td>
</tr>
</tbody>
</table>

Each of the four key ideas about cells has between 4 and 25 related alternative conceptions (AAAS Project 2061, n.d.). Table 5 describes the seven most common middle school alternative conceptions reported by AAAS (n.d.).
Table 5. Seven common middle school alternative concepts related to cells (AAAS, n.d.).

<table>
<thead>
<tr>
<th>Key idea</th>
<th>Alternative conception</th>
<th>Percent frequency of selecting alternative conception</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cells of living organisms do not make molecules for their own growth and repair.</td>
<td>71%</td>
</tr>
<tr>
<td>2</td>
<td>Animal cells do not eliminate their own wastes.</td>
<td>44%</td>
</tr>
<tr>
<td>1</td>
<td>All cells are the same size and shape, i.e. there is a generic cell (AAAS Project 2061, n.d.).</td>
<td>43%</td>
</tr>
<tr>
<td>2</td>
<td>Bacteria do not make molecules for their own growth, extract energy from food, or eliminate their own wastes.</td>
<td>41%</td>
</tr>
<tr>
<td>4</td>
<td>Red blood cells do not supply oxygen to cells of the digestive tract.</td>
<td>40%</td>
</tr>
<tr>
<td>1</td>
<td>There are no single-celled organisms (AAAS Project 2061, n.d.).</td>
<td>39%</td>
</tr>
<tr>
<td>4</td>
<td>The leaves of plants cannot develop from a single fertilized cell.</td>
<td>38%</td>
</tr>
</tbody>
</table>

Other cell related alternative conceptions are often associated with cell typology. Refer to the following examples: (i) there are only eukaryotic cells; (ii) there are two types of cells-plant and animal cells; and (ii) a bacterium is not a cell or a single-celled organism (NGSS Lead States, 2013).

Human Body Systems. The human body systems topic of life science is comprised of four sub-ideas (AAAS Project 2061, n.d.). Refer to Table 6 for a description of these
sub-areas along with their corresponding Next Generation Science Standards (NGSS Lead States, 2013).

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Description</th>
<th>NGSS Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Oxygen, carbon dioxide, and molecules from food are carried to or from cells of the body by means of the circulatory system.</td>
<td>MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
<tr>
<td>2</td>
<td>Most of the carbohydrates, fats, and proteins from the food humans eat must be broken down into smaller molecules before they can enter cells to be used for energy and building materials.</td>
<td>MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
<tr>
<td>3</td>
<td>Lungs take in oxygen molecules and eliminate carbon dioxide.</td>
<td>MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
<tr>
<td>4</td>
<td>Oxygen, carbon dioxide, and molecules from food are carried to or from cells of the body by means for the circulatory system; and molecules from food are broken down into smaller molecules in the digestive tract and then enter the circulatory system by way of capillaries located in the lining of the digestive tract.</td>
<td>MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
</tbody>
</table>
Each of the four key ideas about human body systems has between 3 and 13 related alternative conceptions (AAAS Project 2061, n.d). Table 7 describes the seven most common alternative conceptions reported by AAAS (n.d.).

Table 7. Seven common middle school alternative concepts related to human body systems (AAAS, n.d.).

<table>
<thead>
<tr>
<th>Key idea</th>
<th>Alternative conception</th>
<th>Percent frequency of selecting alternative conception</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Air is distributed through the body in air tubes (Arnudin &amp; Mintzes, 1985; Catherall, 1982).</td>
<td>53%</td>
</tr>
<tr>
<td>4</td>
<td>Molecules from food are distributed by way of special tubes, not by way of the circulatory system, to the rest of the body (Catherall, 1982).</td>
<td>53%</td>
</tr>
<tr>
<td>2</td>
<td>Simple sugars have to be broken down into smaller molecules before they can enter the cells of the body.</td>
<td>52%</td>
</tr>
<tr>
<td>3</td>
<td>If food could not be digested, it would stay in the body and would not be eliminated (AAAS Project 2061, n.d.).</td>
<td>42%</td>
</tr>
<tr>
<td>2</td>
<td>Fatty acids have to be broken down into smaller molecules before they can enter the cells of the body.</td>
<td>41%</td>
</tr>
<tr>
<td>3</td>
<td>The heart is the mixing place for air and blood (Catherall, 1982).</td>
<td>38%</td>
</tr>
<tr>
<td>1</td>
<td>Blood does not carry simple sugar molecules to the cells of the body.</td>
<td>38%</td>
</tr>
</tbody>
</table>
Additional human body system alternative conceptions have been identified within the concept of cellular respiration. For example: (i) glucose is produced as a result of respiration (Al Khawaldeh, & Al Olaimat, 2010); (ii) there is no mechanistic connection between metabolism and respiration (Anderson, Sheldon, & Dubay, 1990); and (iii) in order for the metabolism to increase, respiration must decrease (Anderson, et al., 1990).

*Interdependence in Ecosystems.* The life science topic of interdependence in ecosystems is comprised of three sub-ideas (AAAS Project 2061, n.d.). Refer to Table 8 for a description of these sub-areas along with their corresponding Next Generation Science Standards (NGSS Lead States, 2013).
### Table 8. Key ideas about interdependence in ecosystems (AAAS Project 2061, n.d.; NGSS Lead States, 2013).

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Description</th>
<th>NGSS Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All organisms, both land-based and aquatic, are connected to other organisms by their need for food. This results in a global network of interconnections, which is referred to as a food web.</td>
<td>MS-LS2-2: Construct an explanation that predicts patterns of interactions among organisms across multiple ecosystems.</td>
</tr>
<tr>
<td>2</td>
<td>In all environments, individual organisms that depend on the same resource may compete for that resource when it is limited. Resources that can be limited include food, space, water, shelter, and light.</td>
<td>MS-LS2-4: Construct an argument supported by empirical evidence that changes to physical or biological components of an ecosystem affect populations. MS-LS2-3: Develop a model to describe the cycling of matter and flow of energy among living and nonliving parts of an ecosystem.</td>
</tr>
<tr>
<td>3</td>
<td>Given adequate resources and an absence of disease or predators, populations of organisms in ecosystems can increase at rapid rates. Finite resources and other factors limit their growth.</td>
<td>MS-LS2-1: Analyze and interpret data to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem. MS-LS2-3: Develop a model to describe the cycling of matter and flow of energy among living and nonliving parts of an ecosystem.</td>
</tr>
</tbody>
</table>
Each of the three key ideas about the interdependence in ecosystems has between two and ten related alternative conceptions according to AAAS Project 2061 (n.d). Table 9 describes the seven most common alternative conceptions reported by AAAS (n.d.).
Table 9. Seven common middle school alternative concepts related to the interdependence in ecosystems (AAAS, n.d.).

<table>
<thead>
<tr>
<th>Key idea</th>
<th>Alternative conception</th>
<th>Percent frequency of selecting alternative conception</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Competition between organisms always involves direct, aggressive interaction. Exploitative competition (e.g., getting to the resource before other organisms) is not competition (AAAS Project 2061, n.d.).</td>
<td>63%</td>
</tr>
<tr>
<td>1</td>
<td>If a population in a food web is disturbed, there will be little or no effect on populations that are not within the linear sequence in the food web (Webb &amp; Boltt, 1990).</td>
<td>31%</td>
</tr>
<tr>
<td>2</td>
<td>Plants do not compete for resources (AAAS Project 2061, n.d.).</td>
<td>31%</td>
</tr>
<tr>
<td>1</td>
<td>Changes in the size of a population of organisms will affect only those populations of organisms that are directly connected to it in a feeding relationship, not organisms that are one or more steps removed/away from it (Griffiths &amp; Grant, 1985; Webb &amp; Boltt, 1990).</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>If a population in a food web is disturbed, there will be little or no effect on populations below it in the food web (e.g. if a predator is removed, no effect on the prey: Webb &amp; Boltt, 1990; Leach, Driver, Scott &amp;Wood-Robinson, 1996).</td>
<td>18%</td>
</tr>
<tr>
<td>3</td>
<td>Organisms higher in a food web eat everything that is lower in the food web (Griffiths &amp; Grant, 1985).</td>
<td>16%</td>
</tr>
<tr>
<td>1</td>
<td>Organisms of the same species do not compete with each other for resources (AAAS Project 2061, n.d.).</td>
<td>15%</td>
</tr>
</tbody>
</table>
The science education enterprise has identified additional alternative conceptions related to the interdependence in ecosystems and these alternative conceptions tend to reference food chains and webs. Several researcher found that students believe: (i) organisms higher in a food web eat everything that is lower in the food web; (ii) a food web can be interpreted as a simple food chain; (iii) the top of the food chain has the most energy because energy accumulates up the food chain, (iv) populations higher on a food web increase in size because they deplete those lower in the food web; and (v) varying the population size of an organism will affect all other organisms to the same degree (Griffiths & Grant 1985; Munson 1991).

The concept of unlimited resources or growth also appears to be difficult for students given that a number of researchers found that students believe the following ideas: (i) populations increase until limits are reached, and then they crash and go extinct (McComas, 2002); (ii) populations display either constant growth or decline (Munson, 1991); and (iii) some ecosystems have limitless resources and provide an opportunity for limitless growth of a population (Munson, 1991).

The literature also suggests there are alternative conceptions linked to competition, including predator/prey traits and behaviors such as (i) plants do not compete for light (NGSS Lead States, 2013); (ii) species coexist in an ecosystem because of their compatible needs and behaviors and they get along (Munson, 1991); and (iii)
carnivores are big and/or ferocious, whereas, herbivores are passive and/or smaller (Gallegos, Jerezano, & Flores, 1994).

*Matter and Energy.* The life science topic of matter and energy in living systems is comprised of four sub-ideas (AAAS Project 2061, n.d.). Refer to Table 10 for a description of these sub-areas along with their corresponding Next Generation Science Standards (NGSS Lead States, 2013).
Table 10. Key ideas about matter and energy in living systems (AAAS Project 2061, n.d.; NGSS Lead States, 2013).

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Description</th>
<th>NGSS Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All organisms need food as a source of molecules that provide chemical energy and building materials.</td>
<td>MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
<tr>
<td>2</td>
<td>Plants make their own food in the form of sugar molecules from carbon dioxide molecules and water molecules. In the process of making sugar molecules, oxygen molecules are produced as well.</td>
<td>MS-LS1-6: Construct a scientific explanation based on evidence for the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms.</td>
</tr>
<tr>
<td>3</td>
<td>All organisms, including plants and animals, have mechanisms for storing molecules from food for later use.</td>
<td>MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
<tr>
<td>4</td>
<td>All organisms need food as a source of molecules that provide chemical energy and building materials. Plants make their own food in the form of sugar molecules from carbon dioxide molecules and water molecules. In the process of making sugar molecules, oxygen molecules are produced as well.</td>
<td>MS-LS1-6: Construct a scientific explanation based on evidence for the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms. MS-LS1-7: Develop a model to describe how food is rearranged through chemical reactions forming new molecules that support growth and/or release energy as this matter moves through an organism.</td>
</tr>
</tbody>
</table>
Each of the four key ideas about matter and energy in living systems has between five and ten related alternative conceptions (AAAS Project 2061, n.d). Table 11 describes the seven most common alternative conceptions reported by AAAS (n.d.).

Table 11. Seven common middle school alternative concepts related to matter and energy in living systems (AAAS, n.d.).

<table>
<thead>
<tr>
<th>Key idea</th>
<th>Alternative conception</th>
<th>Percent frequency of selecting alternative conception</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Substances in soil are food for plants (Kuech, Zogg, Zeeman, &amp; Johnson, 2003; Leach et al., 1992; Simpson &amp; Arnold, 1982; Stavy, Eisen, &amp; Yaakobi, 1987; Tamir, 1989; Wandersee, 1983).</td>
<td>72%</td>
</tr>
<tr>
<td>1</td>
<td>Water is food for plants (Horizon, n.d.; Lee &amp; Diong, 1999; Vaz, Carola, &amp; Neto, 1997, Wandersee, 1983).</td>
<td>69%</td>
</tr>
<tr>
<td>3</td>
<td>Molecules from food are not stored in the bulbs of plants.</td>
<td>53%</td>
</tr>
<tr>
<td>2</td>
<td>Plants have multiple food sources, not just the sugars they make from water and carbon dioxide (Anderson et al., 1990; Roth &amp; Anderson, 1987).</td>
<td>40%</td>
</tr>
<tr>
<td>1</td>
<td>Water is food for animals (Horizon, n.d.; Lee &amp; Diong).</td>
<td>38%</td>
</tr>
<tr>
<td>2</td>
<td>Food enters a plant through the roots (Anderson et al., 1990; Roth &amp; Anderson, 1987; Simpson &amp; Arnold, 1982; Vaz et al., 1997; Wandersee, 1983).</td>
<td>36%</td>
</tr>
<tr>
<td>3</td>
<td>Molecules from food are not stored in the fat tissue of animals.</td>
<td>33%</td>
</tr>
</tbody>
</table>

In the area of matter and energy, the additional alternative conceptions that have been identified by science education research tend to be associated with plants. For
example: (i) carbon dioxide is absorbed through the roots of plants (Simpson & Arnold, 1982); (ii) the substances that plants absorb from the earth form raw sap, which moves through the stalk and which allows the plant to grow and carry out its other vital functions (Canal, 1999); and (iii) plants make sugars from minerals and water (NGSS Lead States, 2013).

Reproduction, Genes, and Heredity. There are five sub-ideas within the life science topic of reproduction, genes, and heredity (AAAS Project 2061, n.d.). Refer to Table 12 for a description of these sub-areas along with their corresponding Next Generation Science Standards (NGSS Lead States, 2013).
Table 12. Key ideas about reproduction, genes, and heredity (AAAS Project 2061, n.d.; NGSS Lead States, 2013).

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Description</th>
<th>NGSS Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Genetic information is encoded in DNA molecules.</td>
<td>MS-LS3-1: Develop and use a model to describe why structural changes to genes (mutations) located on chromosomes may affect proteins and may result in harmful, beneficial, or neutral effects to the structure and function of the organism.</td>
</tr>
<tr>
<td>2</td>
<td>Every body cell of an individual organism (with a few exceptions) contains an identical set of DNA molecules and, therefore, contains identical genetic information.</td>
<td>MS-LS3-2: Develop and use a model to describe why asexual reproduction results in offspring with identical genetic information and sexual reproduction results in offspring with genetic variation.</td>
</tr>
<tr>
<td>3</td>
<td>Genetic information in the form of DNA molecules is transferred from parents to offspring during reproduction.</td>
<td>MS-LS3-2: Develop and use a model to describe why asexual reproduction results in offspring with identical genetic information and sexual reproduction results in offspring with genetic variation.</td>
</tr>
<tr>
<td>4</td>
<td>DNA molecules provide the cells with instructions for assembling protein molecules from amino acids.</td>
<td>MS-LS3-1: Develop and use a model to describe why structural changes to genes (mutations) located on chromosomes may affect proteins and may result in harmful, beneficial, or neutral effects to the structure and function of the organism.</td>
</tr>
<tr>
<td>5</td>
<td>The protein molecules an organism makes affect the organism’s physical traits, physiology, and behaviors.</td>
<td>MS-LS3-1: Develop and use a model to describe why structural changes to genes (mutations) located on chromosomes may affect proteins and may result in harmful, beneficial, or neutral effects to the structure and function of the organism.</td>
</tr>
</tbody>
</table>
While there are between 13 and 27 alternative conceptions related to reproduction, genes, and heredity, Table 13 describes the seven most common alternative conceptions reported by AAAS Project 2061 (n.d.).

Table 13. Seven common middle school alternative concepts related to reproduction, genes, and heredity (AAAS, n.d.).

<table>
<thead>
<tr>
<th>Key idea</th>
<th>Alternative conception</th>
<th>Percent frequency of selecting alternative conception</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Genes are traits (Marbach-Ad, 2001).</td>
<td>64%</td>
</tr>
<tr>
<td>1</td>
<td>A gene and the expression of the gene as a characteristic or trait are the same thing (Lewis &amp; Kattman, 2004).</td>
<td>64%</td>
</tr>
<tr>
<td>2</td>
<td>The different cell type (skin, muscle, cartilage, etc.) found in a given individual’s body contains different DNA (Hackling &amp; Treagust, 1984).</td>
<td>61%</td>
</tr>
<tr>
<td>1</td>
<td>There are some types of organisms that do not have DNA (Lewis &amp; Wood-Robinson, 2000; Banet &amp; Ayuso, 2000).</td>
<td>56%</td>
</tr>
<tr>
<td>5</td>
<td>The actions of protein molecules do not affect an organism’s behaviors (AAAS Project 2061, n.d.).</td>
<td>48%</td>
</tr>
<tr>
<td>1</td>
<td>The information in the DNA molecules of a plant does not affect the behaviors of the plant (AAAS Project 2061, n.d.).</td>
<td>46%</td>
</tr>
<tr>
<td>3</td>
<td>In sexually reproducing organisms, half of the organism’s body cells contain DNA from the mother and half contain DNA from the father.</td>
<td>45%</td>
</tr>
</tbody>
</table>
Additional science education literature identifies alternative conceptions related to mutations and dominant alleles. The following examples represent alternative conceptions associated with mutations: (i) all mutations are harmful; and (ii) once a mutation is discovered, it can be fixed (Genetics Generation, 2015). Alternative conceptions linked to dominant alleles include the following examples: (i) dominant alleles increase in frequency in a population (Allchin, 2000; Christensen, 2000); (ii) heterozygotes have a selective advantage over other genotypes (Abraham et al., 2014); and (iii) dominant traits are the most common traits in a population (Genetics Generation, 2015).

Curriculum and Instructional Strategies Targeting Alternative Conceptions

Curriculum and instructional strategies that attempt to modify alternative conceptions are often grounded in the theoretical framework of scientific constructivism. Scientific constructivism places an emphasis on the individual’s construction of knowledge (Battista, 2001). This focus on the individual learner “makes it possible to study student’s knowledge construction in ways that are genuinely relevant to teaching” (Cobb, Wood, & Yackel, 1990; Steffe & Kieren, 1994). The constructs of abstraction, reflection, accommodation, and perturbations are fundamental to scientific constructivism. According to Battista (2001, p. 11), abstraction is a “process by which the mind selects, coordinates, combines, and registers in memory a collection of mental items or acts that appear in the attentional field.” Reflection is described as the
“conscious process of mentally replaying experiences, actions, or mental processes and considering their results or how they are composed (p. 11).” Accommodation occurs when a learner’s cognitive structure adapts as a result of a response to new experiences. Perturbations cause disequilibrium and prompt the act of accommodation. Refer to Figure 1 for a diagrammatic overview of the theoretical framework related to scientific constructivism.

*Figure 1. Overview of the theoretical framework of scientific constructivism.*
Assessments Targeting Alternative Conceptions

Utilizing assessments that target students’ alternative conceptions can help guide instruction and curriculum including individualized perturbation events for learners. The strategic use of assessment provides a pathway for two-way communication between the instructor and student. In turn, this communication increases feedback opportunities, which are essential to support student conceptual understanding of science topics (Wilson & Scalise, 2006). According to the National Research Council (Pellegrino, Chudowsky, & Glaser, 2001), there is a deficiency of quality feedback and communication in education as a result of the ineffective utilization of assessment. For example, the National Research Council (Pellegrino et al., 2001, p. 87) suggests, a major law of skill acquisition involves knowledge of results. Individuals acquire a skill much more rapidly if they receive feedback about the correctness of what they have done. If incorrect, they need to know the nature of their mistake. It was demonstrated long ago that practice without feedback produces little learning (Thorndike, 1931). One of the persistent dilemmas in education is that students spend time practicing incorrect skills with little or no feedback. Furthermore, the feedback they receive is often neither timely nor informative. For the less capable student, unguided practice can be practice in doing tasks incorrectly.
Three Key Pedagogical Characteristics of Assessment. Under this overarching umbrella of assessment, there are three key forms of assessment: diagnostic, formative, and summative assessment. Concept inventories can be used to target all three of these key characteristics of assessment. Concept inventories help an instructor measure students’ conceptual understanding by examining learners’ naïve, alternative, existing, and intuitive conceptions (Bransford, Brown, & Cocking, 2000; Kuhn, Amsel, & O’Loughlin, 1988). Furthermore, concept inventories help instructors understand how new knowledge is acquired with respect to students’ prior knowledge through a pre-/post-test administration of the measurement instrument (Marbach-Ad et al., 2010).

Diagnostic assessment, also referred to as pre-assessments, can be utilized to measure learners’ naïve, alternative, existing, and intuitive conceptions before a learning activity is implemented in a course (Glynn, & Duit, 1995). By investigating how students think about concepts prior to engaging in formal learning activities, an instructor can design lesson plans that support individual learners through the use of differentiation. A concept inventory could be deployed at the beginning of a course or before each unit that is embedded in the course in order to measure and analyze where students’ initial conceptual understanding is so that the instructor can then develop an action plan to help scaffold learning (Marbach-Ad et al., 2010).

Formative assessment occurs during the learning activity and can be formal and informal in nature (Bell, 2002). Not only does formative assessment help guide the
teacher’s instruction, but it allows students the opportunity to reflect on their learning as well. Refer to Figure 2 for practices often associated with quality formative assessment (Black & William, 2010).

A summative assessment, also referred to as a post-assessment, can be utilized to measure shifts in learners’ naïve, alternative, existing, and intuitive conceptions after a learning activity has occurred in a course (Glynn & Duit, 1995). A concept inventory could serve as a summative assessment at the end of a course or end of each unit in order to measure and analyze conceptual understanding shifts in the class as a whole, sub-

![Figure 2. Quality formative assessment practices (Black & William, 2010).](image-url)
populations within the class, and/or individual students (D'Avanzo, 2008). Furthermore, because each of the distractor choices on a concept inventory are typically aligned with naïve, alternative, existing, and intuitive conceptions, an instructor could analyze the summative assessment results to find trends in how students think about specific concepts (Marbach-Ad et al., 2010). That is, if the results suggest little growth on a particular concept, then the instructor may need to restructure the learning activities for future courses and/or reteach the concept with the learners who struggled.

**Overview of Life Science Concept Inventories**

This section will highlight life science concept inventories that have been utilized to measure students’ alternative conceptions. Refer to Table 14 for a summary of these concept inventories.
**Table 14. Summary of Life Science Concept Inventories.**

<table>
<thead>
<tr>
<th>Instructional Level</th>
<th>Instrument Name</th>
<th>Conceptual Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tertiary</td>
<td>The Biological Experimental Design Concept Inventory (BEDCI)</td>
<td>Experimental Design</td>
</tr>
<tr>
<td></td>
<td>Diagnostic Question Clusters on Energy and Matter (DQCs)</td>
<td>Energy and Matter</td>
</tr>
<tr>
<td></td>
<td>-The Concept Inventory of Natural Selection (CINS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-The Evolutionary Development Biology Concept Inventory (EvoDevoCI)</td>
<td>Evolution and Natural Selection</td>
</tr>
<tr>
<td></td>
<td>-Measure of Understanding of Macroevolution (MUM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Basic Tree Thinking Assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Developmental Biology Content Survey</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Genetics Literacy Assessment Instrument (GLAI)</td>
<td>Inheritance</td>
</tr>
<tr>
<td></td>
<td>-Meiosis Concept Inventory (Meiosis CI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Host-Pathogen Interactions (HPI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Introductory Molecular and Cell Biology Assessment (IMCA)</td>
<td>Molecular Biology</td>
</tr>
<tr>
<td></td>
<td>-Molecular Life Sciences (MLS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Biology Concept Inventory (BCI)</td>
<td>Biology</td>
</tr>
<tr>
<td></td>
<td>-Internal Transport in Plants and the Human Circulatory Systems</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Breathing and Respiration</td>
<td>Energy and Matter</td>
</tr>
<tr>
<td></td>
<td>-Photosynthesis and Respiration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Covalent Bonding and Photosynthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Diffusion and Osmosis Diagnostic Test (DODT)/ Osmosis and Diffusion Conceptual Assessment (ODCA)</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>Flowering Plant Growth and Development</td>
<td>Growth and Development</td>
</tr>
<tr>
<td></td>
<td>The Generalized Acceptance of Evolution Concept Inventory (GAENE)</td>
<td>Evolution and Natural Selection</td>
</tr>
<tr>
<td></td>
<td>Genetics Literacy</td>
<td>Inheritance</td>
</tr>
<tr>
<td></td>
<td>Secondary-Biology Concept Inventory (S-BCI)</td>
<td>Biology</td>
</tr>
</tbody>
</table>
Tertiary-Level (College-Level) Concept Inventories. This section describes life science concept inventories developed for use in college courses (see Table 15 for a summary).
Table 15. Summary of Tertiary-Level Life Science Concept Inventories.

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>Number of Items</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Biological Experimental Design Concept Inventory (BEDCI)</td>
<td>14</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Osmosis and Diffusion Conceptual Assessment (ODCA)</td>
<td>18</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Diagnostic Question Clusters on Energy and Matter (DQCs)</td>
<td>15</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>The Concept Inventory of Natural Selection (CINS)</td>
<td>20</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>The Evolutionary Development Biology Concept Inventory (EvoDevoCI)</td>
<td>11</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Measure of Understanding of Macroevolution (MUM)</td>
<td>27/1</td>
<td>single-tiered, selected response / open response</td>
</tr>
<tr>
<td>Basic Tree Thinking Assessment</td>
<td>10</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Developmental Biology Content Survey</td>
<td>15</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Genetics Literacy Assessment Instrument (GLAI)</td>
<td>31</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Meiosis Concept Inventory (Meiosis CI)</td>
<td>17</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Host-Pathogen Interactions (HPI)</td>
<td>18</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Introductory Molecular and Cell Biology Assessment (IMCA)</td>
<td>24</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Molecular Life Sciences (MLS)</td>
<td>26</td>
<td>single-tiered, selected response</td>
</tr>
<tr>
<td>Biology Concept Inventory (BCI)</td>
<td>25</td>
<td>single-tiered, selected response</td>
</tr>
</tbody>
</table>
Experimental Design. The Biological Experimental Design Concept Inventory (BEDCI) was developed by Deane, Jeffery, Pollock, and Birol (2014) for the purpose of measuring students’ non-expert like thinking in experiential design in the context of biology (p. 540; p. 549). The target population for this concept inventory is undergraduate-level biology students (pp. 543-544). Each of the BEDCI’s 14 items displays a single-tiered, multiple choice design (p. 952).

Evolution and Natural Selection. The Concept Inventory of Natural Selection (CINS) was developed by Anderson, Fisher, and Norman (2002) for the purpose of measuring students’ understanding of not only the process of natural selection, but also students’ understanding of ecology and genetics (p. 955). This integrative concept approach was employed because knowledge of ecology and genetics serve as the foundation for understanding the concept of natural selection (p. 955). The target population for this assessment is students enrolled in tertiary education programs (p. 953). Each of the CINS’ 20 items displays a single-tiered, multiple choice design (p. 952) and is associated with one of the three item contextual passages (p. 953). The item contextual passages represent authentic evolutionary processes investigated by scientists (p. 953) and include the following contexts: Galapagos finches, Venezuelan guppies, and Canary Island lizards (pp. 971-975).

The Evolutionary Development Biology Concept Inventory (EvoDevoCI) was develop by Perez, Hiatt, Davis, Trujillo, French, Terry, and Price (2013) for the purpose
of measuring tertiary life science majors’ evolutionary development understanding. The target population for this assessment is tertiary biology majors (Perez et al., 2013, p. 673). Each of the EvoDevoCI’s 11 items displays a single-tiered, selected response design (p. 665) and is associated with one of the four item contextual passages (p. 953). The item contextual passages involve the evolutionary development of the following organisms: crayfish, centipedes, minnows, and lizards (p. 670).

The Measure of Understanding of Macroevolution (MUM) was developed by Nadelson, and Southerland (2010) for the purpose of measuring the following five major facets of macroevolution: (i) deep time; (ii) phylogenetic; (iii) speciation; (iv) fossils; and (v) the nature of science (pp. 157-160). The intended population for the MUM is undergraduate college students (p. 162). The MUM is comprised of 27 selective-response items and one open-response item (p. 151).

The Basic Tree Thinking Assessment was developed by Baum, Smith, and Donovan (2005) to measure the principle of common ancestry through the use of a ten selected response items (p. 979). These item displayed different phylogenetic trees (p. 980). This assessment’s target population was not identified.

The Developmental Biology Content Survey was developed by Knight and Wood (2005 to measure content knowledge related to animal development. This survey is comprised of 15 selected response items which are intended for use with college-aged
students who have completed courses in genetics, molecular biology, and cell biology (p. 299).

*Energy and Matter.* The Diagnostic Question Clusters on Energy and Matter (DQCs) instrument was developed by Wilson et al., (2007) and targets college-level biology students’ conceptual understanding of the conservation of matter in biology systems including tracing matter through photosynthesis and cellular respiration (p. 323). This instrument contains 16 total items arranged into clusters comprised of six to eight questions. The items display both multiple-choice and open response formats.

The Osmosis and Diffusion Conceptual Assessment (ODCA) was developed by Fisher, Williams, and Lineback (2011) and includes 18 total items intended for students enrolled in lower- and upper-level undergraduate biology courses. Both instruments target conceptual understanding related to (i) the process of diffusion; (ii) the particulate and random nature of matter; and (iii) the process of osmosis (Fisher et al., 2011; Odom & Barrow, 1995).

*Inheritance.* The Genetics Literacy Assessment Instrument (GLAI) was developed by Bowling, Acra, Wang, Myers, Dean, Markle, Moskalik, and Huether (2008) in order to measure the following six sub-concepts of genetics literacy: (i) nature of the genetic material; (ii) transmission; (iii) gene expression; (iv) gene regulation; (v) evolution; (vi) genetics and society (p. 19). Undergraduate students in introductory
biology genetics courses are the target population of the GLAI (p. 15). The GLAI is comprised of 31 single-tiered selected response items (p. 15).

The Meiosis Concept Inventory (Meiosis CI) was developed by Kalas, O’Neill, Pollock, and Birol (2013) and measures seven constructs including (i) timing of events during meiosis; (ii) gamete formation; chromosomal representation of genotypes; and (iii) relationships between chromosomes, DNA, and chromatids, and the relation to DNA replication (p. 659). The Meiosis CI is comprised of 17 single-tier items that are intended for students enrolled in introductory biology and genetics courses (p. 655).

*Molecular Biology.* The Introductory Molecular and Cell Biology Assessment (IMCA) was developed by Shi, Wood, Martin, Guild, Vicens, and Knight (2010) and includes 24 multiple choice items. Undergraduate students enrolled in introductory molecular and cell biology courses are the target population for this instrument (p. 453). The IMCA includes a variety of ‘learning goals’ including concepts related to evolution, macromolecules, thermodynamics and kinetic characteristics of biochemical reactions, and gene expression (p. 455).

The Molecular Life Sciences (MLS) was developed by Wright and Hamilton (2008) to measure students enrolled in introductory biochemistry and molecular biology courses (p. 216). The MLS is comprised of 26 single-tiered selected response items and targets the following ten ‘big ideas’ associated with equilibrium in molecular life sciences: molecular evolution; information and communication, self-assembly,
compartmentalization, regulation, catalyst, energy and organization, aqueous environment, and complexity of molecular structure (p. 219).

The Host-Pathogen Interactions (HPI) instrument was developed by Marbach-Ad et al. (2009) to measure microbiology majors’ understanding of 13 ‘big ideas’ including the concepts that (i) immune response memory is specific; (ii) microbes adapt and respond to the environment by altering their metabolism; and (iii) microbes have various strategies to cause disease (p. 45). The HPI displays a two-tiered multiple choice design with a total of 18 items (p. 48).

*Comprehensive Biology Assessments.* The Biology Concept Inventory (BCI) was developed by Klymkowsky, Underwood, and Garvin-Doxas (2010) for the purpose of measuring six constructs: (i) diffusion and drift; (ii) energetics and interactions; (iii) molecular properties and functions; (iv) genetic behaviors; (v) evolutionary processes; and (vi) experimental design (p. 7). The BCI contains 30 selected response items developed for undergraduate students enrolled in molecular level biology courses (p. 1).

*Secondary-Level Concept Inventories.* This section describe life science concept inventories developed for use in high school and middle school classrooms (see Table 16 for a summary).
Table 16. Summary of Secondary-Level Life Science Concept Inventories.

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>Number of Items</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Transport in Plants and the Human Circulatory Systems</td>
<td>28</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Breathing and Respiration</td>
<td>12</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Photosynthesis and Respiration</td>
<td>13/1</td>
<td>two-tiered, selected response / open response</td>
</tr>
<tr>
<td>Covalent Bonding and Photosynthesis</td>
<td>15</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Diffusion and Osmosis Diagnostic Test (DODT)</td>
<td>12</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Flowering Plant Growth and Development</td>
<td>26</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>The Generalized Acceptance of Evolution Concept Inventory (GAENE)</td>
<td>14</td>
<td>single-tiered, Likert-scale</td>
</tr>
<tr>
<td>Genetics Literacy</td>
<td>26</td>
<td>two-tiered, selected response</td>
</tr>
<tr>
<td>Secondary-Biology Concept Inventory (S-BCI)</td>
<td>25</td>
<td>single-tiered, selected response</td>
</tr>
</tbody>
</table>

*Evolution and Natural Selection.* The Generalized Acceptance of Evolution Concept Inventory (GAENE) was developed by Smith, Snyder, and Devereaux (2016) for the purpose of measuring student acceptance of evolution (p. 1292). That is, this measurement instrument does not target evolution related conceptual understanding or
religious belief systems as it centers on whether a learner is willing to accept the concept of evolution (p. 1289). The target student population for this assessment is secondary and college students enrolled in introductory science courses (p. 1289). The GAENE’s 14 items employ a single-tiered, Likert-scale (1-5) design (p. 1298). An example of one of the GAENE’s items is “I would be willing to argue in favor of evolution in a small group of friends (p. 1304).”

**Inheritance.** The Genetics Literacy concept inventory was developed by Tsui and Treagust (2009) to measure scientific reasoning in the context of Mendelian genetics (i.e. monohybrid crosses and pedigrees) (p. 1073). This concept inventory displays a two-tiered item design with a total of 26 questions (p. 1073). The target population for this assessment are students in grades 10 to 12 (p. 1073).

**Growth and Development.** The Flowering Plant Growth and Development was developed by Lin (2004) and displays are two-tiered item format with a total of 26 questions (p. 175). The following constructs are included in this concept inventory: (i) flowering plants life cycle and reproduction; (ii) seed germination; (iii) plant nutrition; and (iv) mechanism of growth and development (pp. 185-186). High school students enrolled in biology courses are the designated population for this instrument (p. 178).

**Energy and Matter.** The Diffusion and Osmosis Diagnostic Test (DODT) was developed by Odom and Barrow (1995) to assess college-level biology majors and non-
majors along with high school aged students. The DODT is comprised of 12 two-tiered selected response items.

The Breathing and Respiration concept inventory was developed by Mann and Treagust (1998) and is comprised of 12 two-tiered items (p. 55). This instrument’s target population is students enrolled in grades 10 and 11. The Breathing and Respiration concept inventory measures students’ understanding of concepts related to breathing, gas exchange, and respiration (p. 55).

The Internal Transport in Plants and the Human Circulatory Systems concept inventory was developed by Wang (2004) and includes the following constructs: (i) functions of internal transports in plants and the human body; (ii) routes of internal transport in plants; (iii) causal mechanisms of internal transport in plants; (iv) routes of the human circulatory system; and (v) causal mechanisms of the human circulatory system (p. 137). This instrument has a two-tiered design and is comprised of 28 total items (p. 131). The target student population for this assessment is elementary and secondary students (p. 141).

The Photosynthesis and Respiration concept inventory was developed by Haslam and Treagust (1987) and includes 13 two-tiered selected response items and one open-ended response question. The Photosynthesis and Respiration concept inventory is intended for use with high school aged students enrolled biology courses.
The Covalent Bonding and Photosynthesis concept inventory was developed by Treagust (1986) for the purpose of measuring high school chemistry students’ alternative conceptions in the following constructs: (i) all organisms, plants and animals, respire continually; (ii) oxygen is taken in during respiration; (iii) photosynthesis takes place only in the presence of light energy; and (iv) oxygen gas is given off by the green leaves (or green stems) during the process of photosynthesis (p. 203). This instrument is comprised of 15 items and display a two-tier design (p. 204).

Comprehensive Biology Assessments. The Secondary-Biology Concept Inventory (S-BCI) measures high school students’ understanding in the following core concepts: (i) evolution and diversity; (ii) population interactions; (iii) growth and reproduction; (iv) inheritance; and (v) energy and matter (Stammen, et al., 2016). The S-BCI is comprised of 25 single-tiered items.

Of the 21 instruments reviewed, 16 assessments were tailored to college-aged students; and only one concept inventory, the Internal Transport in Plants and the Human Circulatory Systems instrument, was developed for use with middle school students (Wang, 2004). This survey of literature also highlights that dearth of life science concept inventories specifically developed for students in middle school science classrooms that can be used over all three middle school grade levels (6 grade through 8 grade). Additionally, the majority of these instruments focus only on specific conceptual topics (i.e. genetics, cell biology, physiology and organismal system, or evolution). The use of
these targeted assessments is impractical to implement in a middle school context because a pre-/post-testing regime for each life science unit of study would require ample in-class instructional time and may lead to testing fatigue. Thus, there is a need within the science education enterprise for a life science concept inventory that is specifically developed for 6th to 8th graders and is easy to administer at each middle school grade level.

*The Middle School-Life Science Concept Inventory (MS-LSCI)*

The Middle School-Life Science Concept Inventory (MS-LSCI) was conceived after the development of the Secondary-Biology Concept Inventory (S-BCI). Originally, the S-BCI was developed for the purpose of measuring the conceptual understanding of secondary grade-level biology students (grades 7 to 12) (Stammen, et al., 2016). However, the S-BCI was validated for use only at the high school-level and not at the middle school-level. This lack of validation became evident when middle-school teachers, as a part of Biology Modeling research workshop and research project, were asked to implement the S-BCI in their classrooms. The results of this initial field-testing of the S-BCI in a middle school context suggested several validity concerns based on the feedback from these middle school teachers. These concerns centered on the S-BCI’s misalignment with middle school standards, which is an indication of a lack of content validity. Furthermore, these teachers suggested that language demands of the S-BCI were too difficult for middle school students. Additionally, one of the teachers stated that
in the future she will not use the S-BCI in her classroom which is indicative of a lack of social validity. Therefore, the inception of MS-LSCI was the result of teacher feedback from the initial field-testing and the absence of validation studies of the S-BCI within the context of middle school.

**Overarching Research Aim**

Within the science education research community, several concept inventories which target narrow biology concepts have been developed. Fewer concept inventories have been developed which attempt to target broader biology concepts covered over a longer period of time (a single course, a grade level, or grade bands). Further, literature suggests that no fully developed concept inventory targeting the major life science concepts covered throughout middle school classrooms (i.e. grades 6 through 8) exists. The purpose of this research is to fill this gap through the development of a valid and reliable measurement instrument, the MS-LSCI.

**Overview of Studies**

Study one focuses on how data from a multi-panel expert review and student interviews were used in the Middle School-Life Science’s (MS-LSCI) item content validation and iterative refinement process. Twelve items were identified as having content validity concerns by the expert panel. Generally, these content validity concerns fell within two categories: (i) imprecise phrasing and (ii) age inappropriateness. A total of 26 items were identified as displaying content validity issues during the student
interviews. These 26 items fell into one of three categories: (i) imprecise phrasing, (ii) contextual ambiguity, and (iii) formatting/diagrammatic complexity. Using the data from the multi-panel expert review and student interviews, the items with content validity concerns were refined and modified before the items were field tested.

The purpose of study two is to describe the MS-LSCI’s validation and item selection process. Specifically, this study focuses on the psychometric functioning of the 60 field-tested MS-LSCI items using Rasch analysis. The results of this development, refinement, and evaluation process suggest that the 25-item MS-LSCI is a valid instrument in that the items appear to be unidimensional, item and person measures display little misfit, and the reliability values suggest replicability within the targeted sample. Based on these results, the MS-LSCI has the potential to help fill the gap in the assessment tools available to measure middle school student life science concepts in conjunction with alternative conceptions.

The purpose of study three is to describe the difference among 6th, 7th, and 8th graders’ performance on the MS-LSCI’s 60 field tested items. The results suggest that MS-LSCI performance was significantly lower for 6th graders relative to 8th graders and that six items can be linked to this difference in performance. These six items fell within the MS-LSCI’s core concepts of evolution and diversity, population interaction, and growth and reproduction. More importantly, these six items shed light on the contextual differences between 6th and 8th grade performance. The contextualized item differences
discover in this study may be attributed to middle school life science curriculum sequencing and the perseverance of specific alternative conceptions.
References


Chapter 2: Study One

The Development and Validation of the Middle School-Life Science Concept Inventory (MS-LSCI): A Qualitative Study of Construct and Content Validity

Abstract

Concept inventories are widely used by researchers and educators as a tool for assessing learner’s content knowledge. A wide variety of concept inventories exist across disciplines, however, most of these concept inventories have been validated for college and high school-level students but not middle school students. To help fill this gap, our goal is to develop the Middle School-Life Science Concept Inventory (MS-LSCI). The purpose of this study is to describe the MS-LSCI’s item development and content validation process. Specifically, this study focuses on how data from a multi-panel expert review and student interviews were used in the MS-LSCI’s item content validation and iterative refinement process. Twelve items were identified as having content validity concerns by the expert panel. Generally, these content validity concerns fell within two categories: (i) imprecise phrasing and (ii) age inappropriateness. A total of 26 items were identified as displaying content validity issues during the student interviews. These 26 items fell into one of three categories: (i) imprecise phrasing, (ii) contextual ambiguity, and (iii) formatting/diagrammatic complexity. Using the data from the multi-panel expert review and student interviews, the items with content validity
concerns were refined and modified before the items were field tested. The next step in the development of the MS-LSCI is the psychometric analysis of validity and reliability using the Rasch model.

Introduction

Science for All Americans’ Project 2061 of the American Association for the Advancement of Science (AAAS Project 2016, n.d.), the National Research Council (2007; 2012), and the Next Generation Science Standards (NGSS Lead States, 2013) have outlined a vision for science education reform founded in standards-based instructional practices. With this vision in mind, middle school (grades 6 to 8) science educators along with all K-12 teachers strive to develop meaningful learning experiences for students working towards mastery in science disciplines. Life science is one of the major science disciplines emphasized in middle school science classrooms (NGSS Lead States, 2013). Middle school life science concepts are misunderstood by many students as evidenced by the litany of alternative conceptions research that exists within science education research enterprise (AAAS Project 2061, n.d.; Understanding Evolution, 2015; NGSS Lead States, 2013; Genetics Generation, 2015; McComas, 2002).

Akin to all teachers, middle school life science educators often seek out targeted instruction that supports learners as they overcome persistent and pervasive alternative conceptions (McComas, 2002; National Research Council, 2012). However, before truly targeted and meaningful learning experiences can be developed to confront
individualized gaps in content knowledge and alternative conceptions, educators must first have a method to measure middle school student understanding and identify the alternative conceptions held by each of their students (National Research Council, 2007; Stern & Ahlgren, 2002). While student interviews provide insight into learners’ conceptual understanding, the time required to conduct interviews with every student in a teacher’s classroom is impractical. Concept inventories offer a time efficient alternative to student interviews (Adam & Wieman, 2010; D’Avanzo, 2008; Garvin-Doxas, Klymkowsky & Elrod, 2007; Hestenes, Well & Swackhammer, 1992; Smith & Tanner, 2010).

While several concept inventories exist to measure life science alternative conceptions, the instruments often contain advanced topics that are outside of the scope of middle school content standards; and thus, are generally deemed inappropriate for use with middle school students [e.g., Biology Concept Inventory (BCI) for undergraduate students (Klymkowsky, Underwood & Garvin-Doxas, 2010); Secondary Biology Concept Inventory (SBCI) for high school students (Stammen, Lan, Schuchaerdt, Malone, Ding, Sabree & Boone, 2016)]. Other concept inventories focus only on specific conceptual topics [e.g. Concept Inventory of Natural Selection (CINS) (Anderson, Fisher & Norman, 2002); EvoDevoCI (Perez, Hiatt, Davis, Trujillo, French, Terry & Price, 2013)]. Given that the middle school life science standards require coverage of multiple
concepts, the utilization of several topic specific concept inventories would lead to the loss of valuable instructional time and the possibility of student testing fatigue.

To help fill this gap, the goal of this study was to develop an easy to administer and analyze concept inventory that targets middle school students’ (grades 6 to 8) understanding of life science concepts. In this paper, we describe the structure and constructs of the Middle School-Life Science Concept Inventory (MS-LSCI) in conjunction with the instrument’s qualitative validation process, which included expert panel reviews and student interviews.

Literature Review

Access to measurement instruments that provide valid insights on students’ alternative conceptions in life science are valued by science educators as diagnostic assessments and by the science education research enterprise as a component to conducting quasi-experimental studies. While there are several instruments developed for college and high school aged students, few instruments exist that target middle school students’ understanding of life science concepts.

*College and High School-Level Life Science Instruments*

Only a few concept inventories have been developed for use with middle school aged students [e.g. Internal Transport in Plants and the Human Circulatory Systems instrument (Wang, 2004)]. Given this dearth of middle school life science concept
inventories, this section will review instruments developed to measure life science concepts at the collegial and high school levels.

With exception to the Biology Concept Inventory (BCI), the vast majority of college and high school-level life science measurement instruments illuminate a narrow breadth of concepts. At the undergraduate level, the Concept Inventory of Natural Selection (CINS), the Conceptual Assessment of Natural Selection (CANS), and Assessing Contextual Reasoning about Natural Selection (ACRONS) were developed to measure the principles of natural selection (Anderson, Fisher & Norman, 2002; Kalinowski, Leonard & Taper, 2016; Nehm, Beggrow, Opfer & Ha, 2012).

The measurement of energy and matter concepts in the context of life science has also been the focus of the science research community. Several of these energy and matter instruments target concepts at the undergraduate level and include (i) The Diagnostic Question Clusters on Energy and Matter (DQCs) (Wilson et al., 2007), (ii) Osmosis and Diffusion Conceptual Assessment (ODCA) (Fisher, Williams & Lineback, 2011), and Diffusion and Osmosis Diagnostic Test (DODT) (Odom & Barrow, 1995). Other energy and matter instruments target concepts at the secondary level: (i) Internal Transport in Plants and the Human Circulatory Systems (Wang, 2004), (ii) Breathing and Respiration instrument (Mann & Treagust, 1998), (iii) Photosynthesis and Respiration instrument (Haslam & Treagust, 1987), and (iv) Covalent Bonding and Photosynthesis instrument (Treagust, 1986).
Other life science concept inventories focus on (i) inheritance [e.g. Genetics Literacy Assessment Instrument (GLAI) (Bowling, Acra, Wang, Myers, Dean, Markle, Moskalik & Huether, 2008) and Genetics Literacy instrument (Tsui & Treagust, 2009)], (ii) molecular biology [e.g. Molecular Life Sciences (MLS); Wright & Hamilton, 2008] and Introductory Molecular and Cell Biology Assessment (IMCA) (Shi, Wood, Martin, Guild, Vicens & Knight (2010)], and (iii) growth and development [e.g. Flowering Plant Growth and Development (Lin, 2004)]. The Flowering Plant Growth and Development inventory (Lin, 2004) was developed for use with high school students while the other instruments associated with inheritance and molecular biology center on concepts explored in college.

Theoretical Framework

*Qualitative Design and Validation Process*

The development and validation of concept inventories often includes the following components: (i) test philosophy, (ii) item development, (iii) item review, and (iv) field testing (Liu, 2010; Schmeiser & Welch, 2006). Assessing content validity is a fundamental step in providing evidence that a concept inventory is a valid measurement instrument (Mussio, & Smith, 1973). Content validity is a measure of how accurate an assessment’s items measures the concepts of a given construct and is established using both qualitative and quantitative methodologies (Liu, 2010). Content validation for this study was examined using a life science expert panel and a middle school teacher expert.
panel along with student interviews. Refer to Figure 3 for additional summary information related to design and validation process for developing a concept inventory. This study will describe the qualitative design process associated with the MS-LSCI’s test philosophy, item development, and item review stages. The quantitative field testing component of the design and validation process of the MS-LSCI will be explicated in chapter 3.

Figure 3. Summary of the design and validation process for a concept inventory.
Research Goals

(1) Establish the content validity of the MS-LSCI’s items qualitatively through expert panel reviews and student interviews.

Methods

The goal in developing the MS-LSCI was to design a measurement instrument that centered on students’ understanding of life science concepts that could provide measures related to the thinking of a large, diverse samples of middle-school aged students (grades 6-8). This concept inventory needed to be shown reliable and valid within the measured middle school population while also displaying the ability to discriminate the abilities of students who hold varying levels of life science conceptual knowledge. Additionally, the MS-LSCI needed to display social validity in that the instrument must be user-friendly for middle school teachers and educational researchers while also being linked to commonly held middle school alternative conceptions and NGSS concepts. With these goals at the forefront of the study, the MS-LSCI items were developed based on alternative conceptions identified from a review of alternative conception research, teacher feedback, and student interviews.
Summary of the design and validation process for the MS-LSCI

A pool of assessment items \( N=60 \) available for field testing was created for consideration of inclusion in the final version of the MS-LSCI. This original item pool consisted of 50 items (10 items per core concept). These items were modified based on feedback from life science experts and middle school teacher experts as well as the insights gained during student interviews. Ten additional items were added to the pool following the multi-panel expert review and student interviews. Refer to Figure 4 for a summary of the MS-LSCI’s design and validation process.

Figure 4. Summary of the design and validation process for the MS-LSCI (Note: * represents when item revision occurred).

Test Philosophy: Instrument Structure and Constructs

The structure and constructs of the MS-LSCI were influenced by the identified needs of a group of life science teachers and experts. This group advocated for an
instrument that could be administered during a single class period (30 to 40 minutes; 20 to 30 items) with a selected response item format that could be easily scored. In addition, this group was interested in an instrument that could shed light on how their students performed on NGSS life science concepts along with the alternative conceptions held by their students.

This group was also surveyed about which concepts represent the fundamental models in the life science curricula at the middle school level. The following five core concepts (CC) emerged from this survey and represent the MS-LSCI constructs (Table 17): (i) evolution and diversity (CC1), (ii) population interactions (CC2), (iii) growth and reproduction (CC3), (iv) inheritance (CC4), and (v) energy and matter (CC5). Essential questions associated with each of the core concepts were developed by the group to explicate the philosophy and constructs of the MS-LSCI’s. In addition, a crosswalk between the core concepts, essential questions, and each of the NGSS life science disciplinary core ideas in middle school was developed (NGSS Lead States, 2013). Both the core concepts and essential questions operationalized the MS-LSCI’s constructs. These constructs represent the foundation of the MS-LSCI, and thus spearheaded item development for the instrument.
Table 17. Essential questions associated with the MS-LSCI core concepts (NGSS Lead States, 2013; Stammen, et al, 2016).

<table>
<thead>
<tr>
<th>Core concepts</th>
<th>Essential questions</th>
<th>NGSS Crosswalk</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC1. Evolution and diversity</td>
<td>How and why do populations change over time?</td>
<td>LS4: Biological evolution: Unity and diversity</td>
</tr>
<tr>
<td>CC2. Population interactions</td>
<td>How and why do populations in a system interact with other populations over time?</td>
<td>LS2: Interactions, energy, and dynamics relationships in ecosystems</td>
</tr>
<tr>
<td>CC3. Growth and reproduction</td>
<td>How is information preserved during reproduction while still producing the variation observed in life?</td>
<td>LS1: From molecules to organisms: Structure and processes</td>
</tr>
<tr>
<td>CC4. Inheritance</td>
<td>How are traits passed from parents to offspring?</td>
<td>LS3: Heredity: Inheritance and variation of traits</td>
</tr>
<tr>
<td>CC5. Energy and matter</td>
<td>How and why do energy and matter transfer within and across systems?</td>
<td>LS1: From molecules to organisms: Structure and processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LS2: Interactions, energy, and dynamics relationships in ecosystems</td>
</tr>
</tbody>
</table>
Item Development

Initially, 10 selective response items were written or adapted from existing instruments targeting each core concept (see Table 18). This original item pool consisted of 50 items. Ten additional items were added to the pool following the multi-panel expert review and student interviews. Thus, a total of 60 single-tiered items were developed for field testing with each item comprised of a question stem and four possible responses. Alternative conceptions identified by empirical research and practitioner observation are represented by the distractor choices associated with each item. Figure 5 is a sample item associated with the population interactions core concept and highlights how the distractor choices are linked to alternative conceptions.

Table 18. Number of items per MS-LSCI core concepts.

<table>
<thead>
<tr>
<th>Core concepts</th>
<th>Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC1. Evolution and diversity</td>
<td>10</td>
</tr>
<tr>
<td>CC2. Population interactions</td>
<td>10</td>
</tr>
<tr>
<td>CC3. Growth and reproduction</td>
<td>10</td>
</tr>
<tr>
<td>CC4. Inheritance</td>
<td>10</td>
</tr>
<tr>
<td>CC5. Energy and matter</td>
<td>10</td>
</tr>
</tbody>
</table>
The following diagram represents a food web within an aquatic ecosystem.

![Food Web Diagram]

Varying the size of the population of the frogs__________.

a. affects the populations that are directly connected to the frog population.
b. affects the entire ecosystem to some degree in varying ways.
c. does not affect the entire ecosystem because some populations are less important.
d. will affect all other populations within the ecosystem to the same degree.

<table>
<thead>
<tr>
<th>Answer Choice</th>
<th>Alternative Conception</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Varying the population size of species will only affect the others that are directly connected through a food chain.</td>
<td>Griffiths &amp; Grant 1985; Munson 1991</td>
</tr>
<tr>
<td>B</td>
<td>Correct response</td>
<td>Stamp, 2015</td>
</tr>
<tr>
<td>C</td>
<td>Some organisms are more valuable to an ecosystem than others.</td>
<td>Teacher observation, N.D.</td>
</tr>
<tr>
<td>D</td>
<td>Varying the population size of an organism will affect all other organisms to the same degree.</td>
<td>Munson, 1991</td>
</tr>
</tbody>
</table>

*Figure 5.* MS-LSCI sample item highlighting the alternative conception distractor choices [item adapted from (Stamp, 2015)].
Item Review

Expert panels. The multi-panel expert critique of the MS-LSCI items involved receiving feedback from two distinct panels: life science experts and middle school teacher experts. Both groups of panelists critically assessed the MS-LSCI items for (i) factual inaccuracies, (ii) unaddressed alternative conceptions, (iii) diagrammatic accuracy, (iv) readability of the sentence structure and item formatting, (v) presence of item bias, and (vi) age-appropriateness.

The life science expert panel consisted of four faculty members who represent four distinct universities across the US. These faculty members had between eight and 20 years of experience in higher education with specializations in biological sciences and/or life science education.

The middle school teacher panel was comprised of five educators representing different public and private middle schools in the Midwest of the US. These panelists had between eight and 24 years of middle childhood science teaching experience and all possessed either a masters or doctoral degree.

Student Interviews. The interview stage the MS-LSCI design and validation process include two age groups of students: undergraduate students and middle school students. Using a think aloud interview structure, both groups of interviewees were asked to explain their understanding of a collection of the items’ question stem and selected response choices. Each student reviewed between 5 and 20 items. The students
were asked to reflect on the areas of the items that were confusing including vocabulary terms, syntax, formatting, and diagrammatic representations.

A total of eight undergraduate students from a large, Midwest, public university and ten middle school students from a large, Midwest, urban private school participated in this stage of the MS-LSCI’s item context review process. Of the ten middle school students, three were 6th graders, four were 7th graders, and the remaining three students were 8th graders.

The undergraduate student interviews occurred before the middle school student interviews for the following reasons: (i) there are fewer barriers in gaining access to undergraduate students (i.e. teacher, parent/guardian, and student consent/assent) and (ii) the undergraduate students served as initial screeners of the items before the middle school student interviews. Additionally, the undergraduate students provided insights on how they would approach each item from a test taking strategy perspective. This initial screening of the items by undergraduate students allowed further edits to be made to the item before the higher-stakes middle school interviews occurred.

Analysis

This analysis will describe how the MS-LSCI items were iteratively reviewed and revised based on data from the multi-panel expert reviews and student interviews (refer to Figure 6 for an overview of this analysis).
Figure 6. Summary of the MS-LSCI’s item review and revision (arrows represent item revision).

**Expert Panel Review**

The first stage in establishing the content validity of the MS-LSCI included an expert panel review. From this review, items that were designated as having ‘content validity concerns’ were modified for student interviews. Of the 50 items reviewed, 12 items were identified by the expert panels as having content validity concerns. Generally, the content validity concerns of these items fell within two categories: (i) age inappropriateness, and (ii) imprecise phrasing. Of the 12 items containing content validity concerns, two questions were labeled as lacking age appropriateness, ten
questions were labeled as lacking precise phasing, and the remaining two questions were labeled as ‘other’. Each of the two foremost categories are briefly defined, and then an exemplar is provided to illustrate the specific content validity concern for each category.

The first category, age inappropriateness, is associated with items that were either too advanced/mature or too simplistic for middle school students. In addition, the expert panel members tended to find these items’ context inappropriate for middle school aged students. The second category, imprecise phrasing, was also identified during the expert panel review. The lack of precise phrasing is related to unclear wording that causes the item to have content inaccuracies and/or has the potential to create distractions for test-takers.

Exemplar I: Age Inappropriateness

The first exemplar assessment item (see Figure 7) represents an example of a question that was identified as having validity concerns during the expert panel review due to a lack of age appropriateness. This question was incorporated into the initial round of testing for the MS-LSCI because the item aligned with the MS-LSCI’s population interactions core concept.
Predation: shark (predator) and seal (prey)  
Parasitism: lamprey (parasite) and fish (host)

What can be said about predation and parasitism?

A. Predation and parasitism are the same because one organism is benefited and the other is harmed.
B. Predation and parasitism are the same because both cause the host to die.
C. Predation is different than parasitism because the prey and predator populations have a long-term relationship.
D. Predation and parasitism are different because predation usually results in death for the prey while in parasitism the host can remain alive.

Figure 7. Original item displaying age inappropriateness.

The original question was reviewed by both the Biology Expert Panel and High School Expert Panel. The expert panel review data suggested that the lack of age appropriateness may cause students to become distracted or induce anxiety. For example, a Middle School Teacher Expert Panel member suggested “the shark picture may induce some feelings of anxiety in middle school students since shark attacks on people have been augmented by the media and the shark is facing the observer. Perhaps consider a picture that is less intimidating to younger students.” Another Middle School Teacher
Expert Panel member stated, “I would have some of my gifted education kids react with deep sensitivity (to these images). The kids may be distracted on this or future questions.” Therefore, this question was edited to be more age appropriate for students and increase content validity by incorporating a different context for the predation example: a lynx and a hare (see Figure 8).

| Predation: lynx (predator) and hare (prey) | Parasitism: lamprey (parasite) and fish (host) |

What can be said about predation and parasitism?

A. Predation and parasitism are the same because one organism is benefited and the other is harmed.
B. Predation and parasitism are the same because both cause the host to die.
C. Predation is different than parasitism because the prey and predator populations have a long-term relationship.
D. Predation and parasitism are different because predation usually results in death for the prey while in parasitism the host can remain alive.

Figure 8. Edited item displaying age inappropriateness (modifications are highlighted).
Exemplar II: Imprecise Phrasing

The second exemplar assessment item represents a question that was identified as having content validity concerns related to a lack of precise phrasing during the expert panel review. The original item (see Figure 9) was developed by the research team for the MS-LSCI because the item aligned with the MS-LSCI’s inheritance core concept.

If a couple has a one-in-four (25%) chance of having a child with an inheritable disease, then__________.

A. the second-born child born will have a reduced chance of inheriting the disease if their firstborn child has the disease.
B. the firstborn child has the highest chance of inheriting the disease when compared to the second-born child.
✓ C. each child born to this couple will have a one-in-four chance of inheriting the disease.
D. The chances of inheriting the disease will depend on the gender of the child.

Figure 9. Original item displaying imprecise phrasing.

The original question (Figure 9) was evaluated by the Expert Panel Review and was found to display content validity concerns due to imprecise phrasing. For example, a Middle School Teacher Expert Panel member stated, “Couple has many meanings for some kids. Today, family units look very different to kids/families. I’d opt for (the phrasing) a male and a female”. The Middle School Teacher Expert Panel member’s comment not only represented how this teacher predicts that some of her students would think about this question, but it also elucidated that the question stem was lacking precise
phrasing. Based on the feedback from both the High School and Biology Expert Panel members, this question (see Figure 10) was edited to reduce student confusion and increase content validity by adding the phrase male and female.

If a **male and female** couple has a one-in-four (25%) chance of having a child with an inheritable disease, then_________.

A. the second-born child born will have a reduced chance of inheriting the disease if their firstborn child has the disease.
B. the firstborn child has the greatest chance of inheriting the disease when compared to the second-born child.
✓ C. each child born to this couple will have a one-in-four chance of inheriting the disease.
D. the chances of inheriting the disease will depend on the gender of the child.

*Figure 10. Edited item displaying imprecise phrasing (modifications are highlighted).*

In addition to modifying items, the multi-expert panel feedback led to the development of two additional items which resulted in 52 items being included in the student interview review process. Both of these items were adapted from original MS-LSCI questions to improve item readability. The first item added was related to the core concept of population interactions (see *Figure 11*) while the second item added centered on the core concept of energy and matter (see *Figure 12*).
### Original Item

<table>
<thead>
<tr>
<th>Slugs→Toads→Birds→Cats</th>
<th>Slugs→Toads→Birds→Cats</th>
</tr>
</thead>
</table>
| In this food chain, cats_____.
  a. eat everything (slugs, toads, and birds) because they are higher in the food chain.
  b. have more energy than slugs because energy adds up to food chain.
  ✓ c. have cells that can harvest energy as efficiently as slugs, toads, and birds.
  d. have a larger size because they can absorb the energy of slugs, toads, and birds. |
| In this food chain, _____.
  a. cats eat everything (slugs, toads, and birds) because they are higher in the food chain.
  b. cats save more energy than slugs because energy adds up to food chain.
  ✓ c. cats eat birds because they are a food source.
  d. cats are larger in size because they can absorb the energy of the other organisms. |

*Figure 11. Original and adapted population interaction items (Stamp, 2015).*

### Original Item

<table>
<thead>
<tr>
<th>In all living organisms, the functions necessary to sustain life________.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. depend on the consumption of a nutritious diet.</td>
</tr>
<tr>
<td>✓ b. are the result of regulated chemical reactions.</td>
</tr>
<tr>
<td>c. are the result of creating biomolecules for cells.</td>
</tr>
<tr>
<td>d. depend on the consumption of a high-energy diet.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In all living organisms, the functions necessary to sustain life________.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. depend on creating energy molecules because food is a source of chemical energy.</td>
</tr>
<tr>
<td>✓ b. depend on food as a source of molecules because they provide building materials and chemical energy.</td>
</tr>
<tr>
<td>c. depend on the intake of a healthy diet because nutrients are sources of building materials.</td>
</tr>
<tr>
<td>d. depend on food absorption because water, air, and minerals are needed from the environment.</td>
</tr>
</tbody>
</table>

*Figure 12. Original and adapted energy and matter items.*
Student Interview Review

The second stage in establishing the content validity of the MS-LSCI included interviews with undergraduate and middle school students, respectively. From this review, items that were designated as having ‘content validity concerns’ were modified following both student groups’ interview. Of the 52 items reviewed, 26 items were identified by the student interviews as having content validity concerns. The content validity concerns of these items fell within three categories: (i) imprecise phrasing, (ii) contextual ambiguity, and (iii) formatting/diagrammatic complexity. Of the 26 items containing content validity concerns, 14 questions were labeled as displaying imprecise phrasing, eight questions were labeled as having contextual ambiguity, and the remaining four questions were labeled as displaying formatting/diagrammatic complexity issues (see Table 19) for a summary of the distribution of items containing content validity concerns by student interview group). Each category is briefly defined, and then an exemplar is provided to illustrate the specific content validity concern category.
Table 19. Distribution of items with content validity concerns by student interview group.

<table>
<thead>
<tr>
<th>Interview groups</th>
<th>Imprecise phrasing</th>
<th>Contextual ambiguity</th>
<th>Formatting/diagrammatic complexity</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate students</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Middle school students</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>8</td>
<td>4</td>
<td>26</td>
</tr>
</tbody>
</table>

Similar to the expert panel reviews, imprecise phrasing was also identified during the student interview phrase of the MS-LSCI’s development. The imprecise phrasing is related to unclear wording that may cause the item to have content inaccuracies and/or has the potential to create distracts for test-takers. The second category, contextual ambiguity, is associated with items missing the details needed to appropriately frame the content being measured by an item. The interviewees tended to find these items vague, and often asked for clarification and additional information to help frame the question stem and/or selected response choices. Finally, items are classified as having formatting/diagrammatic complexity issues when the interviewees indicated that the item’s structure and/or diagram were difficult to interpret.

Exemplar III: Imprecise Phrasing

This exemplar item represents a question that was identified as having content validity concerns related to a lack of precise phrasing during both the undergraduate and
middle school student interviews. This example item (see Figure 13) is aligned with the MS-LSCI’s population interactions core concept and targets alternative conceptions related to interdependent relationships in ecosystems.

<table>
<thead>
<tr>
<th>Competition for food, space, water, shelter, and light</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Involves only direct, aggressive interaction between individual organisms.</td>
<td></td>
</tr>
<tr>
<td>✔️ B. Involves individual organisms that depend on the same resources.</td>
<td></td>
</tr>
<tr>
<td>C. Does not affect plant populations since they do not need many resources.</td>
<td></td>
</tr>
<tr>
<td>D. Does not occur between organisms in the same population.</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 13. Original item displaying imprecise phrasing.*

The item in Figure 13 as evaluated by both interviewee populations and was found to display content validity concerns due to imprecise phrasing. For example, an undergraduate student stated, “I don't think plants can have aggressive interactions.” A different undergraduate student asked if the term “aggressive could be changed to direct contact.” These interviewees’ comments suggested that they would avoid selecting distractor choice A due to the use of the term aggressive. Based on the insight from this undergraduate student, this question was edited for the middle school student interviews by changing the term *aggressive* to the term *physical* in order to make the distractor choice more appealing to students who hold the alternative conception that competition between organisms always involves direct, physical interactions (see Figure 15) (AAAS Project 2016, n.d.).
Competition for food, space, water, shelter, and light_____________.
   A. Involves only direct, physical interaction between individual organisms.
   ✔️ B. Involves individual organisms that depend on the same resources.
   C. Does not affect plant populations since they do not need many resources.
   D. Does not occur between organisms in the same population.

*Figure 14.* Post-undergraduate student interviews edited item displaying imprecise phrasing.

During the middle school student interviews, a student identified another area where this item could display improvement in phrasing by stating “what about organisms about the bottom of the sea?” This question indicates that the item asks students to generalize about how all organisms on Earth compete for resources including organisms classified as chemotrophs, autotrophs, and heterotrophs. Therefore, the middle school interviewee’s questions suggest that the term light in the question stem should be optional when considering responses that include chemotrophs. Thus, the ‘and/or’ phrase was added to the question stem of this item so that all organisms, regardless of trophology, are taken into account. Figure 15 represents how the item was changed based on the feedback obtained during the middle school level interviews.

Competition for food, space, water, shelter, and/or light_____________.
   A. Involves only direct, physical interaction between individual organisms.
   ✔️ B. Involves individual organisms that depend on the same resources.
   C. Does not affect plant populations since they do not need many resources.
   D. Does not occur between organisms in the same population.

*Figure 15.* Post-middle school interviews edited item displaying imprecise phrasing (modifications are highlighted).
Exemplar IV: Contextual Ambiguity

This exemplar item represents a question that was identified as having content validity concerns related contextual ambiguity during both the undergraduate and middle school student interviews. This example item (see Figure 16) is aligned with the MS-LSCI’s evolution and diversity core concept and targets alternative conceptions related to the inheritance of traits.
On an island with dense grass, you find the following population of mice. Initially, these mice had few predators.

![100 Big Mice](image)

**100 Big Mice**

Stronger than small mice and find it hard to move quickly through dense grass

![100 Small Mice](image)

**100 Small Mice**

Can move quickly through dense grass

On the same island with the big and small mice population, you find 5 small mice who lost their tails when they escaped from predators. These mice can move through the thick grass even faster than other small mice with tails. Does this change your prediction about the island’s mouse population over 1000 generations?

A. Yes, because tailless mice needed to change to be better able to survive.
B. Yes, because tailless mice are better able to survive and will pass the tailless trait down to more offspring.
C. No, because tailless mice might be better able to survive but they can’t pass the tailless trait down to their offspring.
D. No, because organisms are equally able to survive and reproduce under most conditions.

✓ C. No, because tailless mice might be better able to survive but they can’t pass the tailless trait down to their offspring.

During the interview stage of the MS-LSCI development, the item in Figure 16 displayed content validity concerns due to contextual ambiguity. During an undergraduate interview a student asked if, “were the tails were ripped off (by the predators)?” Another undergraduate student stated “I am not seeing the correlation between the changes in the new mouse population. I need more information about the
predator.” Figure 17 represents how the item was changed based on the feedback obtained during the middle school level interviews.

On an island with dense grass, you find the following population of big and small mice. Initially, these mice had few predators.

Another animal population arrives on the island and is a predator to mice. This new animal can move fast in open areas. However, this new animal moves slowly through the dense grass on the island. You find 5 small mice who lost their tails when they escaped from these predators. These tailless mice can move through the thick grass even faster than other small mice with tails. This new animal equally prefers eating both big and small mice. The tailless mice

A. needed to change to be better able to survive.
B. are better able to survive and will pass the tailless trait down to more offspring.
✓ C. might be better able to survive but they cannot pass the tailless trait down to their offspring.
D. are equally able to survive and reproduce.

Figure 17. Post-undergraduate student interviews edited item displaying contextual ambiguity (modifications are highlighted).

The middle school interviews highlighted similar issues with the item. A student asked, “does the new animal eat big and small mice and can the big mice defend
themselves?” This student also felt like there was too much text to read to understand what the question was asking. The interviewees’ comments suggested that additional context was needed to describe the new animal that arrived on the island along with how the tailless mice came to lose their tails. The question was edited by providing an observation data table to help improve readability (Figure 18). Within this data table, additional information is provided about the newly introduced predators along with how the tailless mice lost their extremity: being forcefully removed by a predator.
On an island with dense grass, you find the following population of big and small mice. Initially, these mice had few predators.

![Image of big and small mice]

### Observation Data Table

- Another animal population arrives on the island and is a predator to mice.
- This new animal can move fast in open areas.
- This new animal moves slowly through the dense grass on the island.
- You find 5 small mice who lost their tails when they were ripped off while escaping from these predators.
- These tailless mice can move through the thick grass even faster than other small mice with tails.
- This new animal equally prefers eating both big and small mice.

The tailless mice______.

A. needed to change to be better able to survive.
B. are better able to survive and will pass the tailless trait down to more offspring.
C. might be better able to survive but they cannot pass the tailless trait down to their offspring.
D. are equally able to survive and reproduce.

*Figure 18. Post-middle school interviews edited item displaying contextual ambiguity (modifications are highlighted).*
Exemplar V: Formatting/Diagrammatic Complexity

This exemplar item represents a question that was identified as having content validity concerns related formatting/diagrammatic complexity during both the undergraduate and middle school student interviews. This example item (see Figure 19) is aligned with the MS-LSCI’s growth and reproduction core concept and targets alternative conceptions related to cell processes.
A cell (that is not currently dividing) contains two different versions of the same genetic material as shown in the cell and table below.

Below are 6 possible products of the cell division of the original cell above. These 6 representations are not currently dividing.

Which of the 6 representations above best represents the genetic contents of the product of the cell division if the division produced a skin cell?

A. A
B. B
C. C
D. D
E. E
F. F

Figure 19. Original item displaying formatting/diagrammatic complexity.
The item in Figure 19 was identified as displaying formatting/diagrammatic complexity in both the middle school and undergraduate student interviews. For example, an undergraduate student stated, “I just matched the shapes and did not need the tables.” Another undergraduate student suggested that two of the choices should be removed because it requires too much time to review six different versions of a cell. This student also suggested reducing the volume of text by removing unneeded information. Figure 20 represents how the item was changed based on the feedback obtained during the undergraduate level interviews.
A cell (that is not currently dividing) contains two different versions of the same genetic material as shown in the cell and table below. This cell undergoes a type of cell division called mitosis which results in the production of body cells.

<table>
<thead>
<tr>
<th>Chromosome 1</th>
<th>Chromosome 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="chromosome1.png" alt="Image" /></td>
<td><img src="chromosome2.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Below are 4 possible products of the cell division of the original cell above. These 4 representations are not currently dividing.

A  B  C  D

Which of the 4 cell choices above best represents the product of the cell division if the division produced a skin cell?

- A. A
- B. B
- C. C
- D. D

*Figure 20. Post-undergraduate student interviews edited item displaying formatting/diagrammatic complexity (modifications are highlighted).*
The middle school students expressed the same concerns with this item. One student stated, “it took time to read all of this stuff.” Using this feedback, the item’s formatting and diagrammatic complexity was modified (see Figure 21). Specifically, the text was consolidated into one contextual passage, the gene variant tables were removed, two distractor choices were removed, and the link between the parent cell and daughter cell was made explicit with the use of arrows.
A cell (that is not currently dividing) contains two different versions of the same genetic material as shown in the cell and table below. This cell undergoes a type of cell division called mitosis which results in the production of body cells.

<table>
<thead>
<tr>
<th>Chromosome 1</th>
<th>Chromosome 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Chromosome 1" /></td>
<td><img src="image2" alt="Chromosome 2" /></td>
</tr>
</tbody>
</table>

Which of the 4 cell choices above best represents the product of the cell division if the division produced a body cell?

A. A  
B. B  
C. C  
D. D

*Figure 21. Post-middle student interviews edited item displaying formatting/diagrammatic complexity (modifications are highlighted).*
In addition to modifying the 26 items identified by the student interviews as having content validity concerns, the students’ feedback lead to the development of eight additional items which resulted in 60 items being included in the field-testing phrase of this study. These items were adapted from original MS-LSCI questions to improve item readability. These newly developed items were related to the following MS-LSCI core concepts: evolution and diversity (N=4), population interactions (N=1), growth and reproduction (N=2), and inheritance (N=1) (see Figure 22 for an example of an inheritance item that was added to the item pool).

<table>
<thead>
<tr>
<th>Original Item</th>
<th>Adapted Item</th>
</tr>
</thead>
</table>
| Mutations that occur in DNA sequences during replication are __________. | During DNA replication, the mutations that occur are__________.
| a. usually harmless to the individual who may inherit the mutation. | a. usually harmless to the individual who may inherit the mutation. |
| b. usually harmful to the individual who may inherit the mutation. | b. usually harmful to the individual who may inherit the mutation. |
| c. always fixed once discovered through cellular processes. | c. always fixed once discovered through cellular processes. |
| d. often related to genetic diseases in offspring. | d. often related to genetic diseases in offspring. |

*Figure 22. Original and adapted inheritance items.*

Results

The primary stages in establishing the content validity of the MS-LSCI’s items included a multi-panel expert review and student interviews (see Figure 23 for an overview of the MS-LSCI’s item review and revision process).
Figure 23. Summary of the MS-LSCI’s item review and revision (arrows represent item revision).

Of the 50 items reviewed by the life science experts and middle school teacher experts, 12 items were identified as having content validity concerns with the concerns generally falling within two categories: (i) age inappropriateness, and (ii) imprecise phrasing. In addition to guiding the item revision process, the multi-expert panel feedback led to the development of two new items which resulted in 52 items being included in the student interview review process. Of the 52 items reviewed, 26 items were identified during the student interviews as having content validity concerns. The content validity concerns emerging from the interviews fell within three categories: (i)
imprecise phrasing, (ii) contextual ambiguity, and (iii) formatting/diagrammatic complexity. Eight additional items were developed to improve item readability which resulted in a total of 60 items in the MS-LSCI item pool. Refer to Table 20 for a summary of the item distribution of the field testing items per the MS-LSCI’s core concepts.

Table 20. Number of items per MS-LSCI core concepts used during field testing.

<table>
<thead>
<tr>
<th>Core concepts</th>
<th>Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC1. Evolution and diversity</td>
<td>13</td>
</tr>
<tr>
<td>CC2. Population interactions</td>
<td>13</td>
</tr>
<tr>
<td>CC3. Growth and reproduction</td>
<td>13</td>
</tr>
<tr>
<td>CC4. Inheritance</td>
<td>10</td>
</tr>
<tr>
<td>CC5. Energy and matter</td>
<td>11</td>
</tr>
</tbody>
</table>

Discussion and Conclusion

Access to measurement instruments that provide valid insights on students’ alternative conceptions in life science are valued by science educators as diagnostic assessments and by the science education research enterprise as a component to conducting quasi-experimental studies. While several concept inventories exist to measure life science alternative conceptions, the instruments often contain advanced topics that are outside of the scope of middle school content standards; and thus, are
generally deemed inappropriate for use with middle school students (e.g., Klymkowsky, et al., 2010; Stammen, et al., 2016) or focus only on specific conceptual topics (e.g., Anderson, et al, 2002; Perez, et al, 2013). To help fill this gap, the overarching goal of our project is to develop a valid, easy to administer and analyze concept inventory that targets middle school students’ (grades 6 to 8) understanding of life science concepts.

Assessing content validity is a fundamental step in providing evidence that a concept inventory is a valid measurement instrument (Mussio, & Smith, 1973). Content validity is related to how accurate an assessment’s items measures the concepts of a given construct and established using both qualitative and quantitative methodologies (Liu, 2010). The goal of this study was to describe the item development process and to establish the content validity of the MS-LSCI’s items qualitatively through expert panel reviews and student interviews. The results of the study elucidated the MS-LSCI items iterative review and revision process associated with qualitative methods utilized while establish the content validity. While the results of this study are a fundamental methodological component of the development and validation of the MS-LSCI, our understanding the quantitative psychometric properties of the instrument also need to be explored in order to corroborate the qualitative findings of this study.

**Future Steps**

The 60 MS-LSCI items were field-tested in to collect the corpus of data necessary to examine the quantitative content validity of the items. The MS-LSCI assessment data
were collected from over 500 middle school students who were in five different schools located in the Midwest of the United States. The Rasch model will be utilized for the psychometric analysis of validity and reliability of the MS-LSCI. Following field testing, this project’s hope is that the MS-LSCI can be used to supplement that growing number of life science measurement instruments.

*Implications*

The abundance of the MS-LSCI’s item requiring modification due to validity concerns discovered by the multi-panel expert reviews and student interviews highlights the importance of a high quality, qualitative review of all assessments include those used in science education research and classrooms. When assessments are utilized without undergoing a thorough qualitative item review, often assumptions related to the accuracy and precision of measuring learning are violated. That is, in the absence of a qualitative review of an instruments’ items, teachers and researchers are more at risk of making invalid inferences about conceptual understanding that influence their students’/subjects’ grades, academic standing, and access to instructional interventions. This study serves as an example of how to conduct a qualitative review of an assessment’s items.

*Limitations*

The MS-LSCI, like all concept inventories, has intrinsic limitations. MS-LSCI needs to be user-friendly for middle school teachers and educational researchers to administer during a typically class period. While the MS-LSCI will target each of the
NGSS life science disciplinary core ideas in middle school, the breadth of concept coverage is restricted given that students need to complete the instrument in a reasonable amount time. Additional, the selected response format, while easy to score, limits the depth at which the instrument can measure students’ conceptual understanding relative to instruments that display open-ended or interview question formats.

One area of improvement for this study centers on sampling. This study’s middle school student interviewee sample included students from one urban school in the US Midwest. For future studies, the middle school student interviews should include students from diverse school settings including school situated in rural and suburban communities. By having a more diverse middle school sample, future studies would be better positioned to discern if item performance is influenced by contextual differences between learning contexts.
References


99


Chapter 3: Study Two

The Development and Validation of the Middle School-Life Science Concept Inventory (MS-LSCI): A Quantitative Study Using Rasch Modeling

Abstract

The purpose of this study is to describe the Middle School-Life Science Concept Inventory’s (MS-LSCI) validation and item selection process. Specifically, this study focuses on the psychometric functioning of the 60 field-tested MS-LSCI items using Rasch analysis. The results of this development, refinement, and evaluation process suggest that the 25-item MS-LSCI is a valid instrument in that the items appear to be unidimensional, item and person measures display little misfit, and the reliability values suggest replicability within the targeted sample. Based on these results, the MS-LSCI has the potential to help fill the gap in the assessment tools available to measure middle school student life science concepts in conjunction with alternative conceptions.

Introduction

Life science is one of the major science disciplines emphasized in middle school science classrooms (NGSS Lead States, 2013). Middle school life science concepts are misunderstood by many students as evidenced by the litany of alternative conceptions research that exists within science education research enterprise (AAAS Project 2061, n.d.; Understanding Evolution, 2015; NGSS Lead States, 2013; Genetics Generation,
Research-based tools called concept inventories are often employed by educators and researchers to measure student understanding and alternative conceptions in science education (Adam & Wieman, 2010; D’Avanzo, 2008; Garvin-Doxas, Klymkowsky & Elrod, 2007; Hestenes, Well & Swackhammer, 1992; Smith & Tanner, 2010). For example, the Covalent Bonding and Photosynthesis concept inventory in biology education (Treagust, 1986) and the Force Concept Inventory (FCI) (Hestenes, et al., 1992) in physics education represent two seminal instruments developed to measure conceptual knowledge.

Several concept inventories exist to measure life science alternative conceptions. However, these instruments are generally deemed inappropriate for use with middle school students because they contain advanced topics that are outside of the scope of middle school content standards. For example, the Biology Concept Inventory (BCI) (Klymkowsky, Underwood & Garvin-Doxas, 2010) covers life science concepts at the undergraduate level and the Secondary Biology Concept Inventory (SBCI) (Stammen, et al., 2016) focuses on life science concepts in high school level biology. Other concept inventories focus only on a narrow set of specific conceptual topics [e.g. Concept Inventory of Natural Selection (CINS) (Anderson, Fisher & Norman, 2002); EvoDevoCI (Perez, Hiatt, Davis, Trujillo, French, Terry & Price, 2013)]. Given that the middle school life science standards require the coverage of multiple concepts, the utilization of these narrow topic specific concept inventories is problematic for several reasons. First,
students would be required to complete several inventories over a short period of time which may lead to test fatigue. Also, valuable instructional time would be lost if a battery of specific concept inventories were administered. The loss of instructional time would likely result in a lack of buy-in from teachers, administrators, and the school community. In summary, a need exists for a concept inventory that appropriately supports middle school life science teaching and learning.

In order to expand the life science concept inventory options for the middle school grades 6 through 8, this research aims to develop a socially valid instrument in that it is both easily administered and easily analyzed. In this chapter, the development, validation, and psychometric functioning of the 60 field-tested Middle School-Life Science Concept Inventory (MS-LSCI) items will be described using Rasch analysis. Additionally, the item selection process for the final 25-item MS-LSCI is described.

Theoretical Framework

Quantitative Design and Validation Process

The development and validation of concept inventories often includes the following components: (i) test philosophy, (ii) item development, (iii) item review, and (iv) field testing (see Figure 24) (Liu, 2010; Schmeiser & Welch, 2006). This study will describe the quantitative design and validation process specifically associated with the Rasch analysis of MS-LSCI’s field testing data.
Figure 24. Summary of the design and validation process for a concept inventory.

Classical test theory is an essential part of the foundation of instrument measurement. Many concept inventories used in science research have employed the classical test theory (i.e. The Biological Experimental Design Concept Inventory (BEDCI) (Deane, Jeffery, Pollock, & Birol, 2014); The Concept Inventory of Natural Selection (CINS) (Anderson, et al., 2002); and The Evolutionary Development Biology Concept Inventory (EvoDevoCI) (Perez, et al., 2013). Although classical test theory is widely used when developing concept inventories, the method displays the following
circular reasoning shortcomings: item measures are person dependent, and persons’ total scores are item dependent (Boone, Staver, & Yale, 2014; Ding, & Beichner, 2009).

The Rasch model can be used to overcome the aforementioned limitation associated with classical test theory by assuming unidimensionality of an instrument’s latent traits and employing logistic regression to measure the probability of correct responses on discrete items (Boone, et al., 2014; Ding, & Beichner, 2009). Thus, item measures and person measures are mutually independent (Boone et al., 2014; Ding, & Beichner, 2009). Moreover, unlike item response theory (IRT), the Rasch model is not altered to fit the data (Boone, et al., 2014).

Research Goal

(1) Evaluate the measurement properties of the Middle School-Life Science Concept Inventory’s (MS-LSCI) item bank using Rasch modeling in order to determine the item selection of the final 25-item MS-LSCI.

Methods

Test Philosophy Framework

The structure and constructs of the MS-LSCI were shaped by a panel of life science teachers and experts who advocated for an instrument that could be administered during a single class period (30 to 40 minutes; 20 to 30 items) with a selected response item format that could be easily scored. These experts were interested in an instrument that could offer insights on how their students performed on NGSS life science concepts.
and what alternative conceptions their students held. This panel also developed a framework of the major themes and questions in life science curriculum for the MS-LSCI which consisted of the following five core concepts (CC) (see Table 21): (i) evolution and diversity (CC1), (ii) population interactions (CC2), (iii) growth and reproduction (CC3), (iv) inheritance (CC4), and (v) energy and matter (CC5).
Table 21. Essential questions associated with the MS-LSCI core concepts (NGSS Lead States, 2013; Stammen, et al., 2016).

<table>
<thead>
<tr>
<th>Core concepts</th>
<th>Essential questions</th>
<th>NGSS Crosswalk</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC1. Evolution and diversity</td>
<td>How and why do populations change over time?</td>
<td>LS4: Biological evolution: Unity and diversity</td>
</tr>
<tr>
<td>CC2. Population interactions</td>
<td>How and why do populations in a system interact with other populations over time?</td>
<td>LS2: Interactions, energy, and dynamics relationships in ecosystems</td>
</tr>
<tr>
<td>CC3. Growth and reproduction</td>
<td>How is information preserved during reproduction while still producing the variation observed in life?</td>
<td>LS1: From molecules to organisms: Structure and processes</td>
</tr>
<tr>
<td>CC4. Inheritance</td>
<td>How are traits passed from parents to offspring?</td>
<td>LS3: Heredity: Inheritance and variation of traits</td>
</tr>
<tr>
<td>CC5. Energy and matter</td>
<td>How and why do energy and matter transfer within and across systems?</td>
<td>LS1: From molecules to organisms: Structure and processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LS2: Interactions, energy, and dynamics relationships in ecosystems</td>
</tr>
</tbody>
</table>

**Item Development**

Ten single-tiered, selective response items were written or adapted from existing instruments targeting each of the five core concept, and thus, the original item pool...
consisted of 50 items. Alternative conceptions identified by empirical research and practitioner observation are represented by most distractor choices (see Figure 25 for an example item).

How do plants make sugars?

a. Plants make sugars by absorbing sap from the ground.
b. Plants make sugars by absorbing minerals and water from the ground.
c. Plants make sugars from multiple food sources including water and carbon dioxide.
d. Plants make sugars from carbon dioxide and water using sunlight.

<table>
<thead>
<tr>
<th>Answer Choice</th>
<th>Alternative Conception</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>The substances that plants absorb from the earth form raw sap, which moves through the stalk and which allows the plant to grow and carry out its other vital functions.</td>
<td>Canal, 1999</td>
</tr>
<tr>
<td>B</td>
<td>Plants make sugars from minerals and water.</td>
<td>NGSS Lead States, 2013</td>
</tr>
<tr>
<td>C</td>
<td>Plants have multiple food sources, not just the sugars they make from water and carbon dioxide.</td>
<td>Anderson et al., 1990; Roth &amp; Anderson, 1987</td>
</tr>
<tr>
<td>D</td>
<td>Correct response</td>
<td></td>
</tr>
</tbody>
</table>

Figure 25. MS-LSCI sample field-tested item highlighting the alternative conception distractor choices.

*Item Review: Expert Panels and Student Interviews*

The item review stage involved an iterative review and revision cycle which included expert panel reviews and student interviews. The multi-panel expert critique of the items included feedback from two distinct panels. The life science expert
panel consisted of four faculty members from four different universities across the US. These faculty members had between eight and 20 years of experience in higher education with specializations in biological sciences and/or life science education. The middle school teacher panel was comprised of five educators representing different public and private middle schools in the Midwest of the US. These panelists had between eight and 24 years of middle childhood science teaching experience and all had earned at least a graduate degree. Both groups of panelists examined the MS-LSCI items for (i) factual inaccuracies, (ii) unaddressed alternative conceptions, (iii) diagrammatic accuracy, (iv) readability of the sentence structure and item formatting, (v) presence of item bias, and (vi) age-appropriateness.

The interview stage included two age groups of students: undergraduate students and middle school students. Both groups of interviewees were asked to explain their understanding of 5 to 20 items’ question stem and selected response choices using a think aloud interview structure. The students were asked to reflect on the areas of the items that were confusing or difficulty including vocabulary terms, syntax, formatting, and diagrammatic representations.

Eight undergraduate students from a large, Midwest, public university participated in interviews. The undergraduate student interviews occurred before the middle school student interviews for the following reasons: (i) there are fewer barriers in gaining access to undergraduate students (i.e. teacher, parent/guardian, and student
consent/assent) and (ii) the undergraduate students served as initial screeners of the items. Additionally, the undergraduate students provided insights on how they would approach each item from a test taking strategy perspective. This initial screening of the items by undergraduate students allowed further edits to be made to the item before the higher-stakes middle school interviews occurred. Ten middle school students from a large, Midwest, urban private school participated in interviews. The middle school student sample included three 6th graders, four 7th graders, and three 8th graders.

Ten additional items were adapted from items in the original item pool based on feedback from the expert panels and student interviews. These items were developed to address issues with readability of the sentence structure and item formatting. Refer to Figure 26 for an example of one of the additional items that was including in the final item pool based on the feedback that the original item’s tables including gene variants were difficult to interpret. In total, 60 items underwent the iterative review and revision cycle prior to item field-testing.
A cell (that is not currently dividing) contains two different versions of the same genetic material as shown in the cell and table below. This cell undergoes a type of cell division called mitosis which results in the production of body cells.

<table>
<thead>
<tr>
<th>Chromosome 1</th>
<th>Chromosome 2</th>
<th>Gene 1: Variant 1</th>
<th>Gene 2: Variant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Below are 6 possible products of the cell division of the original cell above. These 6 representations are not currently dividing.

A cell (that is not currently dividing) contains two different versions of the same genetic material as shown in the cell and table below. This cell undergoes a type of cell division called mitosis which results in the production of body cells.

Which of the 4 cell choices above best represents the product of the cell division if the division produced a body cell?

- a. A
- b. B
- c. C
- d. D

\[\text{Figure 26. Example of a MS-LSCI field-testing additional item.}\]
Field-testing Data Collection

The 60 MS-LSCI items were distributed between two field testing forms (i.e. Form A and Form B) so that the necessary data were collected to conduct a Rasch analysis of the measurement properties of the MS-LSCI. These two forms shared 15 common items which served as anchor items. Three anchor items were selected for each of the MS-LSCI’s core concepts. Along with the 15 shared items, Form A contained 24 unique items and Form B contained 21 unique items. The inclusion of these anchor items allowed all 60 items to be calibrated on a single unidimensional logit measurement scale. That is, the anchor items included within each form allowed person and item measures to be computed along the same scale regardless of which test form was completed by respondents (Boone et al., 2014). A minimum value of 20 percent of the total items should be linking items in each test form (Liu, 2010). For the MS-LSCI’s field testing, the linking items amounted to 25 percent of the total items in each test form, and thus the percentage of linking items in this study was above the suggested threshold.

The items were administered to 623 middle school students in May 2017. This sample was comprised of 319 6th graders, 188 7th graders, 85 8th graders, and 31 students with a non-identified grade level. These students attended public (N=3) and private (N=1) schools located in the US Midwest. Using a web-based survey tool called Qualtrics, each student completed one of the MS-LSCI’s forms which was randomly assigned by each of the seven teachers who administered the instrument. Teachers
allocated one instructional period or approximately 45 minutes for the completion of the form.

*Rasch Analysis*

The Rasch model (Rasch, 1960) is widely accepted as a preferred psychometric analysis framework for instrument development (Bond & Fox, 2009; Wright & Masters, 1982; Boone et al., 2014). The Rasch model recognizes that selected response item data are discrete and nonlinear which offers an improvement to the traditional instrument development frameworks (i.e. classical test theory) (Wright & Masters, 1982). A Rasch analysis also takes into consideration that items display varying degrees of difficulty while also recognizing that a meaningful instrument must display the ability to measure a single defined trait on a logit scale (i.e. unidimensionality) (Wright & Masters, 1982; Boone, et al., 2014).

Rasch modeling has been applied in a widespread array of research fields (e.g. science education research (Boone, Townsend, & Staver, 2011; Liu, 2010), mathematics education research (Callingham & Bond, 2006), and medical research (Christensen, Kreiner, & Mesbah, 2013). There are a series of Rasch analysis techniques utilized when developing a new instrument like the MS-LSCI. Specifically, Rasch analysis was employed to (i) guide the selection of the MS-LSCI’s items, and (ii) evaluate the psychometric functioning of the instrument (Rasch, 1960; Wright & Stone, 1979; Wright & Masters, 1982; Boone et al., 2014). The succeeding sections will provide an overview 115
of the Rasch analysis techniques used to evaluate the MS-LSCI’s item bank along with a
description of the final item selection of the instrument. Specifically, a Rasch software
program called Winsteps was used to conduct a dichotomous Rasch analysis (Linacre,
2013).

Results

The following series of Rasch analysis techniques were used to measure the
psychometric properties of the MS-LSCI: (i) local independence and unidimensionality
assumption examination, (ii) item difficulty and person ability measure review, and (iii)
Wright map analysis. These analyses were then utilized during the selection of the final
25 items for the MS-LSCI.

Local Independence and Unidimensionality

Two assumptions associated with the Rasch model are that the items are locally
independent and the measurement scale displays unidimensionality. If the items are
locally independent, then the success and/or failure of an item does not affect the success
and/or failure of another item (Crocker & Algina, 2008). The MS-LSCI’s single-tier item
structure helps avoid issues of violating the assumption of local independence since each
item is (i) scored individually, (ii) not linked to another item (two-tiered item structure),
and (iii) not sharing a contextual passage (background description proceeding the
question stem) with another item.
Unidimensionality is the psychometric property that an instrument’s scale measures a single latent trait (a theoretical and unobservable characteristic) (Crocker & Algina, 2008; Allen & Yen, 1979). The MS-LSCI’s latent trait can be defined as conceptual life science knowledge. An evaluation of the MS-LSCI’s unidimensionality involves the investigation of trends in the data that do not fall under the umbrella of this latent trait. The Rasch model’s Item Fit statistics examine how well the items and person response patterns fit the model’s expectations and unidimensionality by measuring infit and outfit metrics (reported in mean-square (MNSQ) values) (Boone et al., 2014). Generally, the MNSQ Item Outfit is inspected in more detail relative to the MNSQ Item Infit because the outfit statistics illuminate items that may be measuring a different latent trait and thus do not fit the model. If an item’s Outfit MNSQ value is greater than 1.3, then the item should be removed from the instrument or replaced with different item (Wright & Linacre, 1994; Boone et al., 2014). All 60 of the MS-LSCI’s items fell below this 1.3 Outfit MNSQ criterion which suggests that the instrument’s items demonstrate acceptable fit along with providing evidence that the MS-LSCI displays unidimensionality (Bond & Fox, 2007).

The examination of the person fit metrics [reported in mean-square (MNSQ) values] helps identify patterns in student answers often hidden in the data set (Boone et al., 2014). That is, person fit statistics highlight students who display idiosyncratic response patterns (i.e. a high performing student who unexpectedly answers an easy item
incorrectly and vice versa). While a person fit value suggests unexpected student response behaviors, it does not expound upon why the response pattern is abnormal, and thus further investigations into the students’ responses are necessary. A general rule of thumb for assessing person fit is to examine if persons Outfit MNSQ values are greater than 1.5 (Wright & Linacre, 1994; Boone et al., 2014). The review of person misfit revealed that 49 students from the sample of N=623 displayed misfit. Thus, approximately eight percent of the sample displayed misfit which is three percent greater than what one would expect by chance. While examining the common characteristics of these misfitting persons, 6th graders tended to have a person Outfit MNSQ value greater than 1.5 when compared to 7th and 8th graders. This pattern may suggest that 6th graders display unexpected student response behaviors by holding more alternative concepts relative to 7th and 8th graders who have more experience engaging in life science learning.

*Item Difficulty and Person Ability*

Unlike other analysis methods, the Rasch model’s item difficulty and person ability are represented on the same logit scale which thereby allows researchers to make comparisons between these two variables (Boone et al., 2014). This side-by-side comparison of item difficulty and person ability is displayed in a Wright map (also known as a person-item map) (Wright & Linacre, 1994). On a Wright map, (i) the left side of figure represents a histogram of person measures and the right side represents item measures, (ii) zero represents items or persons of average ability or difficulty, (iii)
positive values indicate items with higher than average difficulty measures or persons with higher than average ability; and (iv) negative values indicate items with lower than average difficulty measures or persons with lower than average ability (Boone et al., 2014).

The Wright map from the 60 MS-LSCI items provide several insights about the psychometric functioning of the instrument (see Figure 27). First, the map displays a relatively symmetric distribution of person ability ranging from approximately -3 logits to +3 logits. A large proportion of students fall between 0 logits and -1 logits which suggests that the items were too difficult for the majority of the sample and/or that sample is performing below their anticipated ability. Moreover, while there are few gaps in item difficulty, the items tend to measure higher performing students with more precision seeing that all of the items, with exception to item 10 (A10), cluster at approximately -1 logits or above. Given that approximately 86 percent of the sample was comprised of 6th and 7th graders, the underrepresentation of 8th graders in this sample is likely a contributing factor in this misalignment between item difficulty and person ability.
Figure 27. Wright map for 60 item MS-LSCI. Wright map for 60 item MS-LSCI. The left side of the map shows the distribution of student ability and the right side shows the distribution of item difficulty. Items are labeled A1-A60. The labels M, S, and T represent the mean value, one standard deviation, and 2 standard deviations for each distribution, respectively. Each ‘#’ represents 6 students and each ‘.’ represents between 1 and 5 students.
**Item Selection for the Final 25-Item MS-LSCI**

The selection of the final 25 items for the MS-LSCI included the review of the item difficulty measures along with the examination of distribution of the items on the Wright map (refer to Figure 28). The goal of the selection process was to choose five items per core concept which displayed a wide-range of item difficulty and represented varying NGSS life science concepts. Once the items were selected, the overall difficulty of each core concept was determined. The order of core concepts from least difficult to most difficult was (i) population interactions, (ii) energy and matter, (iii) evolution and diversity, (iv) growth and reproduction, and (v) inheritance. Next, the items were ordered from least difficult to most difficult within each core concept. Refer to Table 22 for the description of the final 25 items for the MS-LSCI.
Figure 28. Wright map for the 25 item MS-LSCI. The left side of the map shows the distribution of student ability and the right side shows the distribution of item difficulty. The labels M, S, and T represent the mean value, one standard deviation, and 2 standard deviations for each distribution, respectively. Each ‘#’ represents 8 students and each ‘.’ represents between 1 and 7 students.
<table>
<thead>
<tr>
<th>Final Item Number on 25-Item MS-LSCI</th>
<th>Trial Item Number on the 60-Item Instrument</th>
<th>Item Difficulty Measure (all 60 items)</th>
<th>Connection to NGSS</th>
<th>Core Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>-.78</td>
<td>LS2.A: Interdependent relationships in ecosystems</td>
<td>Population Interactions</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>-.58</td>
<td>LS2.B: Ecosystems dynamics, functioning, and resilience</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>-.76</td>
<td>LS2.A: Interdependent relationships in ecosystems</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>-.25</td>
<td>LS2.C: Interdependent relationships in ecosystems</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>-.11</td>
<td>LS2.B: Ecosystems dynamics, functioning, and resilience</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>-.07</td>
<td>LS1.C: Organization for matter and energy flow in organism</td>
<td>Energy and Matter</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>-.01</td>
<td>LS1.C: Organization for matter and energy flow in organism</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>.01</td>
<td>LS1.C: Organization for matter and energy flow in organism</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>.08</td>
<td>LS1.C: Organization for matter and energy flow in organism</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>49</td>
<td>.23</td>
<td>LS1.C: Organization for matter and energy flow in organism</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>59</td>
<td>-.87</td>
<td>LS3.A: Inheritance of traits</td>
<td>Evolution and Diversity</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>-.59</td>
<td>LS1.B: Growth and development</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>-.49</td>
<td>LS4.A: Evidence of common ancestry and diversity</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>54</td>
<td>-.42</td>
<td>LS4.C: Adaptation</td>
<td></td>
</tr>
<tr>
<td>Final Item Number on 25-Item MS-LSCI</td>
<td>Trial Item Number on the 60-Item Instrument</td>
<td>Item Difficulty Measure (all 60 items)</td>
<td>Connection to NGSS</td>
<td>Core Concept</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------------------------------------------</td>
<td>--------------------------------------</td>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>.36</td>
<td>LS3.B: Variation of traits</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>27</td>
<td>-.95</td>
<td>LS1.A: Structure and function</td>
<td>Growth and Reproduction</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
<td>-.75</td>
<td>LS1.B: Growth and development of organisms</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>30</td>
<td>.43</td>
<td>LS1.B: Growth and development of organisms</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>57</td>
<td>.46</td>
<td>LS1.A: Structure and function</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>56</td>
<td>.94</td>
<td>LS1.A: Structure and function</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>40</td>
<td>-.48</td>
<td>LS3.A: Inheritance of traits</td>
<td>Inheritance</td>
</tr>
<tr>
<td>22</td>
<td>33</td>
<td>-.32</td>
<td>LS3.B: Variation of traits</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>31</td>
<td>.13</td>
<td>LS1.B: Growth and development</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>39</td>
<td>.58</td>
<td>LS3.A: Inheritance of traits</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>35</td>
<td>.91</td>
<td>LS3.A: Inheritance of traits</td>
<td></td>
</tr>
</tbody>
</table>

**Reliability**

The item reliability measure for the 25-item MS-LSCI was .96 and the item separation index was 4.93 which both indicate strong confidence that the item placement will replicate across other samples. The mean item MNSQ was 1.02 which is an indication that the data fit the expected Rasch model expectation of one with approximately two percent of unexplained variance. The person reliability measure was .47 and the person separation index was .95 suggesting that the 25-item MS-LSCI has a
limited capacity to discriminate the sample into two ability levels (i.e. higher performers and low performers) (Boone et al., 2014).

These reliability values for the 25-item instrument were lower than the values associated with the 60-item instrument (item reliability=.97; item separation index=5.43; person reliability=.68; person separation=1.45). This decrease in the reliability values is particularly noticeable in the person reliably measure and person separation index and may be attributed to the reduction in the sample’s ability variance and test length (60 item instrument to the 25 item instrument). Furthermore, better targeted-sampling (i.e. a more even distribution of 6th, 7th, and 8th graders) may improve the person reliability measure and should be considered when field-testing the 25-item MS-LSCI in a future study.

Discussion and Conclusion

Access to measurement instruments that provide valid insights on students’ alternative conceptions in middle school life science are valued by science educators and the science education research enterprise not only as diagnostic and formative assessments, but also as important components of quasi-experiment research projects. Several concept inventories exist to measure life science alternative conceptions, however, these instruments have a tendency of being too advanced for middle schoolers and/or focusing on narrow topics which would require the administration of several assessments within the short period of time.
The goal of this study was to evaluate the measurement properties of the Middle School-Life Science Concept Inventory’s (MS-LSCI) item bank using Rasch modeling in order to determine the item selection of the final 25-item MS-LSCI. The instrument was development by designing items linked to NGSS life science standards with distractor choices associated with alternative conceptions. The items were qualitatively evaluated using a series of expert panels and student interviews, and then the items were field-tested with over 600 middle school students. The psychometric properties of the items were examined using the Rasch model which also was utilized to guide the selection of the 25 items for the final MS-LSCI. The results of this development, refinement, and evaluation process suggest that the 25-item MS-LSCI is a valid instrument in that the items appear to be unidimensional, item and person measures display little misfit, and the reliability values suggest replicability within the targeted sample. Based on these results, the MS-LSCI has the potential to help fill the gap in the assessment tools available to measure middle school student life science concepts in conjunction with alternative conceptions.

**Limitations**

The MS-LSCI, like all concept inventories, has intrinsic limitations. MS-LSCI needed to be user-friendly for middle school teachers and educational researchers to administer during a typically class period. Thus, the MS-LSCI breadth of concept coverage is restricted given that the final instrument consists of 25 items. Additionally,
the selected response format, while easy to score, limits the depth at which the instrument can measure students’ conceptual understanding relative to instruments that display open-ended question formats.

From a psychometric perspective, a maximally valid measurement instrument should examine a single conceptual construct and thus, display unidimensionality. The utility of the MS-LSCI is derived from its ability to measure a diverse set of life science core concepts, all underpinned by life science’ unifying theory of evolution. These life science core concepts are targeted by NGSS standard each year at various middle school grade levels. Thus, rather than measuring a narrow concept exhaustively, the MS-LSCI aims to measure a variety of fundamental life science core concepts that are introduced to students at various points in their time in middle school.

Implications

When used as a diagnostic assessment, the MS-LSCI can serve as a focal point of lesson planning that differentiate instruction based on student conceptual understanding and alternative conceptions. In additional, middle school teacher teams can have targeted discussions about student progress before and after the instruction of the life science units. These discussions would allow teachers the ability to identify areas of strength, in addition to detecting areas of growth, which would allow for the development of instructional goals and benchmarks to monitor areas of improvement. In addition to being a tool to assess students, the MS-LSCI items can be adapted to open-response
questions which could be used during classroom activities and/or incorporated into questionnaires and think aloud interview prompts for research projects. Finally, the unaltered MS-LSCI items have the potential to drive in-class discussions including argumentative based discourse. For example, students may engage in evidence based argumentation when exploring and explaining which responses represent alternative conceptions. In the context of science education research, the MS-LSCI has the potential for being used in quasi-experimental studies. For example, the MS-LSCI can be employed to compare different instructional approaches by examining the difference between students’ pre- and post-test results. Research centered on grade-level comparisons of life science conceptual understanding may also benefit from the utilization of the MS-LSCI.

*Future Research*

In order to further examine the MS-LSCI’s validity including generalizability, future research is needed to replicate and extend the findings in this study in a wide variety of middle school contexts and student samples. Therefore, additional field-testing of the 25-item MS-LSCI is needed.
References


Chapter 4: Study Three

A Contextualized Item Comparison Study of Middle Schooler Performance on the Middle School-Life Science Concept Inventory (MS-LSCI)

Abstract

The purpose of this study is to describe the difference among 6th, 7th, and 8th graders’ performance on Middle School-Life Science Concept Inventory’s (MS-LSCI) 60 field tested items. The results suggest that MS-LSCI performance was significantly lower for 6th graders relative to 8th graders and that six items can be linked to this difference in performance. These six items fell within the MS-LSCI’s core concepts of evolution and diversity, population interaction, and growth and reproduction. More importantly, these six items shed light on the contextual differences between 6th and 8th grade performance. The contextualized item differences discovered in this study may be attributed to middle school life science curriculum sequencing and the pervasiveness of specific alternative conceptions.

Introduction

Life science is a major science discipline emphasized in all middle school science grade levels (NGSS Lead States, 2013). Recognized daily by teachers and evidenced by the litany of alternative conceptions research that exists within science education research enterprise, concepts associated with middle school life science curriculum are often
misunderstood by students (AAAS Project 2061, n.d.; Understanding Evolution, 2015; NGSS Lead States, 2013; Genetics Generation, 2015; McComas, 2002). Concept inventories are often used by educators and researchers to measure student understanding and alternative conceptions in science education (Adam & Wieman, 2010; D’Avanzo, 2008; Garvin-Doxas, Klymkowsky & Elrod, 2007; Hestenes, Well & Swackhammer, 1992; Smith & Tanner, 2010). The Force Concept Inventory (FCI) (Hestenes, Well & Swackhammer, 1992) in physics education and The Covalent Bonding and Photosynthesis concept inventory in biology education (Treagust, 1986) represent two pioneering instruments developed to measure students’ conceptual knowledge.

Several concept inventories exist to measure life science alternative conceptions. The Biology Concept Inventory (BCI) (Klymkowsky, Underwood & Garvin-Doxas, 2010) covers life science concepts at the undergraduate level, whereas the Secondary Biology Concept Inventory (SBCI) (Stammen, Lan, Schuchaerdt, Malone, Ding, Sabree & Boone, 2016) focuses on life science concepts in high school level biology. Other concept inventory focus on a narrow set of specific conceptual topics and tend to measure undergraduate level concepts [e.g. Concept Inventory of Natural Selection (CINS) (Anderson, Fisher & Norman, 2002); EvoDevoCI (Perez, Hiatt, Davis, Trujillo, French, Terry & Price, 2013)]. While these concept inventories are readily available, these instruments are generally deemed inappropriate for use with middle school students.
because they contain advanced topics that are outside of the scope of middle school content standards.

To help fill this gap, the Middle School-Life Science Concept Inventory (MS-LSCI) was developed that targets middle school students’ (grades 6 to 8) understanding of life science concepts. During the MS-LSCI’s development and validation, 60 items were field tested with over 600 middle school students in order to determine which 25 items would comprise the final version of the MS-LSCI. In this paper, we explore the differences between 6th, 7th, and 8th graders’ performance on the MS-LSCI’s 60 field tested items. In addition, a Wright map is used to explore which MS-LSCI items each middle school grade levels display a higher probability of correctly answering between middle school grade levels.

Research Goals

(1) Is there a significant difference among 6th, 7th, and 8th graders’ performance on the MS-LSCI?

(2) If a significant difference is measured, then what MS-LSCI items are linked to the difference in performance among grade levels?

Methods

Data Collection Instrument Format

The MS-LSCI’s item pool consisted of 60 items. To reduce internal validity threats by way of test fatigue, the 60 items were distributed between two field testing
forms (i.e. Form A and Form B). Each form shared 15 common items which are often used as anchor items for psychometric testing including Rasch analysis. Form A contained 24 unique items with 39 items in total, and Form B contained 21 unique items with 36 items in total.

Sample

The MS-LSCI test forms were administered to 623 middle school students in May 2017. This sample was comprised of 319 6th graders, 188 7th graders, 85 8th graders, and 31 students with a non-identified grade level. These students attended public (N=3) and parochial (N=1) schools located in the US Midwest. The 31 students with an unspecified grade level were removed from the sample for this study. Following the investigation of significant outliers, two 7th graders and one 6th who scored 0 percent and 2.7 percent were identified as unusual and were removed from the sample. Additionally, 45 students were removed from the sample given that these students did not reach all of the items on their test form. Thus, the final sample included 544 students with 294 6th graders, 169 7th graders, and 81 8th graders. In this sample, the schools’ middle school life science instruction was underpinned by either state developed model curriculum or state adopted Next Generation Science Standards (NGSS) (NGSS Lead States, 2013).

Data Collection Procedures

Using a web-based survey tool called Qualtrics, each student completed one of the MS-LSCI forms that was randomly assigned by each of the seven teacher who
administered the instrument. Teachers allocated one instructional period or approximately 45 minutes for the completion of the form. In addition, both forms requested demographic information from the student participants and included items such as students’ grade level.

**Measures**

As this study aims to explore whether or not there is a significant difference among 6th, 7th, and 8th graders’ performance on the MS-LSCI, the following two measures were examined using analysis of variance (ANOVA): grade level and performance score.

**Grade level**

The data associated with the independent variable, titled grade level, was collected from the demographic information provided by students during the data collection phase of the study. In the Qualtrics forms, students selected one of the following options to represent their grade level: 6th, 7th, or 8th. These grade level categories were translated into the following respective codes within a data spreadsheet: 1, 2, and 3.

**Performance score**

The performance score measure represents the dependent variable in this study. Person measures (in logits) represent the performance scores for this study and were derived from the Rasch analysis of the MS-LSCI’s item pool.
Results

A one-way analysis of variance (ANOVA) was utilized to determine if a statistically significant difference exists between 6th, 7th, and 8th graders’ performance on the MS-LSCI. Then, a Wright map produced from the Rasch model is used to explore which MS-LSCI items are linked to the difference among these grade levels.

The assumptions associated with one-way analysis of variance were met prior to conducting the analysis. The dependent variable (performance score) and independent variable (grade level) are measured at the interval level and categorical level, respectively. The data displays independence of observations and the independent variable has mutually exclusive and exhaustive categories. The data display no significant outliers after the removal of the three students who scored 0 percent and 2.7 percent. The $p$-values for the Shapiro-Wilk test for the 6th grade student groups is 0.001 suggesting that the data are not normally distributed. However, the histogram suggests that the data are approximately normal. In addition, the Shapiro-Wilk test $p$-values of .071 and .090 for the 7th and 8th grade student groups suggest the data for these groups are normally distributed. The assumption of a homogeneity of variances was found to be violated according to the Levene’s test ($F = 5.32, p = .005$). Therefore, the Welch one-way ANOVA along with the Fishers Least Significant Difference (LSD) post-hoc test were utilized.
ANOVA

A Welch one-way ANOVA was conducted to compare the mean MS-LSCI performance scores between 6th, 7th, and 8th graders. The results of the analysis revealed a statistically significant difference in MS-LSCI performance score means for the three grade levels \(F(2, 541) = 2.985, p = .016\). The Fishers Least Significant Difference (LSD) post-hoc test indicates that MS-LSCI performance was significantly lower for 6th graders relative to 8th graders \((p=.03)\). MS-LSCI performance did not differ significantly between 7th graders and the other grade levels. Refer to Table 23 for the logit mean MS-LSCI performance scores per grade level.

Table 23. MS-LSCI logit mean performance scores per grade level.

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>N</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>6th Grade</td>
<td>294</td>
<td>-.86*</td>
<td>.033</td>
</tr>
<tr>
<td>7th Grade</td>
<td>169</td>
<td>-.76</td>
<td>.043</td>
</tr>
<tr>
<td>8th Grade</td>
<td>81</td>
<td>-.69*</td>
<td>.079</td>
</tr>
<tr>
<td>Total</td>
<td>544</td>
<td>-.80</td>
<td>.025</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level.

Wright Map

Rasch analysis techniques were used to measure the psychometric properties of the MS-LSCI. Item difficulty and person ability can be represented on the same logit scale which thereby allows researchers to make comparisons between these two variables (Boone, Stayer & Yale, 2014). This side-by-side comparison of item difficulty and
person ability is displayed in a Wright map (also known as a person-item map) (Wright & Linacre, 1994). On a Wright map, (i) the left side of figure represents a histogram of person measures and the right side represents item measures, (ii) zero represents items or persons of average ability or difficulty, (iii) positive values indicate items with higher than average difficultly measures or persons with higher than average ability; and (iv) negative values indicate items with lower than average difficultly measures or persons with lower than average ability (Boone et al., 2014).

In addition, Wright maps have the power to bring contextualized meaning to the statistics related measurement instruments like the MS-LSCI (Boone et al., 2014). Namely, group comparison means can be plotted on a Wright map providing insights on which items a group is more likely to answer correctly. For this study, the 6th, 7th, and 8th grade MS-LSCI performance means were plotted on Wright map associated with the field-testing of the 60 MS-LSCI items.

First, the grade level logit means (see Table 23) were plotted on the 60 MS-LSCI item Wright map (see Figure 29). The contextualized item differences excluded the 7th grade group given that the grade level means were significant only between the 6th grade group and 8th grade group. The horizontal lines for 6th and 8th grade means mark the boundary between the items that 8th graders had a greater than a 50 percent probability of correctly answering (all items below the 8th grader mean line) and those items that 6th graders had less than a 50 percent probability of correctly answering (all items below the
6\textsuperscript{th} grader mean line) (Boone, Stayer & Yale, 2014). In addition, 8\textsuperscript{th} graders displayed a higher probability of correctly answering items 3, 4, 15, 19, 20, and 25 relative to 6\textsuperscript{th} graders. Finally, using Winsteps’ Table 20.1, the grade level performance means were calculated (Table 24).

\textit{Table 24.} Grade level performance means to logit mean conversion.

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>N</th>
<th>Percentage Mean</th>
<th>Logit Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>6\textsuperscript{th} Grade</td>
<td>294</td>
<td>31.7*</td>
<td>-.86*</td>
</tr>
<tr>
<td>7\textsuperscript{th} Grade</td>
<td>169</td>
<td>33.3</td>
<td>-.76</td>
</tr>
<tr>
<td>8\textsuperscript{th} Grade</td>
<td>81</td>
<td>35.0*</td>
<td>-.69*</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level.
Figure 29. Wright map for 60 item MS-LSCI with the 6th, 7th, and 8th grade level logits means are plotted at -.86, -.76, and -.69 logits, respectively. The left side of the map shows the distribution of student ability and the right side shows the distribution of item difficulty. Items are labeled A1-A60. The labels M, S, and T represent the mean value, one standard deviation, and 2 standard deviations for each distribution, respectively. Each ‘#’ represents 5 students and each ‘.’ represents between 1 and 4 students.
**Item Context**

Based on the ANOVA results and Wright map with plotted grade level means, the following items represent the difference in 6th and 8th grader performance on the MS-LSCI: 3, 4, 15, 19, 20, and 25. These items will be examined to identify the contextual differences in performance in between 6th and 8th graders.

The six items identified as being the contextual differences between 6th and 8th grader performance were related to the MS-LSCI core concepts of (i) evolution and diversity, (ii) population interactions, and (iii) growth and reproduction (see Table 25). In relation to NGSS, the items were linked to the following Life Science Disciplinary Core Ideas: (i) From Molecules to Organisms: Structures and Processes (LS1), (ii) Interactions, Energy, and Dynamic Relationships in Ecosystems (LS2), and (iii) Biological Evolution: Unity and Diversity (LS4) (NGSS Lead States, 2013).

*Table 25.* Crosswalk between MS-LSCI items, NGSS Disciplinary Core Ideas, and MS-LSCI core concepts.

<table>
<thead>
<tr>
<th>MS-LSCI Item</th>
<th>NGSS Life Science Sub-Idea</th>
<th>MS-LSCI Core Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 4</td>
<td>LS4: Biological Evolution: Unity and Diversity</td>
<td>Evolution and Diversity</td>
</tr>
<tr>
<td>15, 19, 20</td>
<td>LS2: Interactions, Energy, and Dynamic Relationships in Ecosystems</td>
<td>Population Interactions</td>
</tr>
<tr>
<td>25</td>
<td>LS1: From Molecules to Organisms: Structures and Processes</td>
<td>Growth and Reproduction</td>
</tr>
</tbody>
</table>
Evolution and Diversity Core Concept Items

MS-LSCI items 3 and 4 are linked to the concepts of natural selection and evidence of common ancestry (NGSS Lead States, 2013). Item 3 targeted students’ understanding of natural selection whereby variations of traits in a population increase some organisms’ chances of survival and reproduction leading to a predominance of such traits in a population (see Figure 30). This item’s distractor choices were related to the alternative conceptions that (i) changes in traits are based on need (Anderson et al., 2002), (ii) individually acquired modifications can be passed on to offspring (Sundberg, 2003), and (iii) all organisms are equally able to survive and reproduce.

When examining the response patterns between grade levels, 6th graders had a higher tendency to select distractor choice B relative to 8th graders which suggests that the 6th graders in this sample may be holding onto Lamarkian alternative conceptions (see Table 26). In comparison to their 6th grade counterparts, the 8th graders in this sample selected distractor choice A which centered on need-based changes in traits. These data suggest that students’ alternative conceptions related to natural selection may shift from a Lamarkian perceptive to a needs-based perspective between grades 6th and 8th. A comparable amount of 6th and 8th graders selected distractor D. This response pattern may suggest that some middles schoolers, regardless of grade level, believe that all members of a population are nearly identical, and therefore are equally able to survive and reproduce (Anderson et al. 2002).
On an island with dense grass, you find the following population of big and small mice. Initially, these mice had few predators.

Another animal population arrives on the island and is a predator to mice. This new animal can move fast in open areas. However, this new animal moves slowly through the dense grass on the island. You find 5 small mice who lost their tails when they were ripped off while escaping from these predators. These tailless mice can move through the thick grass even faster than other small mice with tails. This new animal equally prefers eating both big and small mice. The tailless mice______.

A. needed to change to be better able to survive.
B. are better able to survive and will pass the tailless trait down to more offspring.
C. might be better able to survive but they cannot pass the tailless trait down to their offspring.
D. are equally able to survive and reproduce.

Figure 30. MS-LSCI item 3.

Table 26. Item 3 response pattern.

<table>
<thead>
<tr>
<th>Response</th>
<th>6th Graders Selection Percentage</th>
<th>8th Graders Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9.9</td>
<td>13.1</td>
</tr>
<tr>
<td>B</td>
<td>28.4</td>
<td>16.7</td>
</tr>
<tr>
<td>C</td>
<td>47.6</td>
<td>56.0</td>
</tr>
<tr>
<td>D</td>
<td>14.1</td>
<td>14.3</td>
</tr>
</tbody>
</table>
MS-LSCI item 4 examined students’ understanding of evidence of common ancestry specifically in relation to chimpanzees and humans (see Figure 31). The following alternative conceptions were also targeted by item 4: (i) species with non-apparent/superficial similarities have no similarities at all (AAAS Project 2061, n.d.), (ii) humans evolved from chimpanzees, and (iii) speciation requires one species to replace the other. The response pattern for item 4 displays similarity between 6th and 8th graders (see Table 27). This similar response pattern may indicate that alternative conceptions related to common ancestry in the context of human and chimpanzees remain persistent between grades. Explicit and targeted curriculum may be needed to support students as they develop their scientific understanding of this concept.

Which statement best describes the common ancestry of humans and chimpanzees?

a. These organisms do not have a common ancestor because there are few physical and genetic similarities between them.

b. These organisms do not have a common ancestor because both species are alive and not extinct.

✓ c. These organisms share a common ancestor because there are many physical and genetic similarities between them.

d. These organisms share a common ancestor because humans evolved from chimpanzees.

Figure 31. MS-LSCI item 4.
Table 27. Item 4 response pattern.

<table>
<thead>
<tr>
<th>Response</th>
<th>6th Graders Selection Percentage</th>
<th>8th Graders Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10.5</td>
<td>9.1</td>
</tr>
<tr>
<td>B</td>
<td>19.0</td>
<td>14.3</td>
</tr>
<tr>
<td>C</td>
<td>51.6</td>
<td>52.1</td>
</tr>
<tr>
<td>D</td>
<td>19.0</td>
<td>24.5</td>
</tr>
</tbody>
</table>

Population Interactions Core Concept Items

MS-LSCI items 15, 19, and 20 are linked to independent relationships in ecosystems (AAAS Project 2061, n.d.). Item 15 targets the concept that organisms/population of organisms compete for the same resources if there are similar resource requirements among these organisms within an ecosystem (see Figure 32). For item 15, the distractor choices related to alternative conceptions are (i) competition between organisms always involves interactions that are direct and aggressive, (ii) non-animal organisms like plants do not compete for resources, and (iii) organisms within a species do not compete with each other for resources (AAAS Project 2061, n.d.). Both 6th and 8th graders in this sample displayed a higher propensity of selecting distractor choice A which is linked to the alternative conception that competition between organisms always involves interactions that are direct and aggressive (see Table 28).
Competition for food, space, water, shelter, and/or light_____________.

a. Involves only direct, physical interaction between individual organisms.
✓ b. Involves individual organisms that depend on the same resources.
c. Does not affect plant populations since they do not need many resources.
d. Does not occur between organisms in the same population.

Figure 32. MS-LSCI item 15.

Table 28. Item 15 response pattern.

<table>
<thead>
<tr>
<th>Response</th>
<th>6th Graders Selection Percentage</th>
<th>8th Graders Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>31.2</td>
<td>23.5</td>
</tr>
<tr>
<td>B</td>
<td>46.5</td>
<td>52.9</td>
</tr>
<tr>
<td>C</td>
<td>15.3</td>
<td>11.8</td>
</tr>
<tr>
<td>D</td>
<td>7.0</td>
<td>11.8</td>
</tr>
</tbody>
</table>

Item 19 deals with generalizing patterns of trophic typologies in that the question stem requires students to compare the needs and behaviors of herbivores and carnivores (see Figure 33). This item includes the following alternative conceptions: (i) herbivores are small and passive organisms, (ii) herbivores are large and aggressive organisms, and (iii) carnivores require more food than herbivores (AAAS Project 2061, n.d.). The 6th graders displayed a higher inclination of selecting distractor choice A which is associated with the alternative conceptions that herbivores are small and passive organisms (see Table 29). The 8th graders had a greater tendency of selecting C which is related to the alternative conception that carnivores require more food than herbivores. This shift in the alternative conception response pattern may indicate that middle schoolers are
moving from a macro-level to a more micro-level understanding of how organisms in
different trophic levels differ.

When comparing carnivores (meat-eaters) to herbivores (plant-eaters), ______________.

a. Carnivores are large and aggressive organisms.
b. Herbivores are small and passive organisms.
c. Carnivores require more food than herbivores.
   ✓ d. Herbivores and carnivores have diverse needs.

Figure 33. MS-LSCI item 19.

Table 29. Item 19 response pattern.

<table>
<thead>
<tr>
<th>Response</th>
<th>6th Graders Selection Percentage</th>
<th>8th Graders Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23.2</td>
<td>17.6</td>
</tr>
<tr>
<td>B</td>
<td>15.9</td>
<td>8.8</td>
</tr>
<tr>
<td>C</td>
<td>15.2</td>
<td>23.5</td>
</tr>
<tr>
<td>D</td>
<td>45.7</td>
<td>50.0</td>
</tr>
</tbody>
</table>

Item 20 targets the various benefits acquired by the mutualistic relationships
interactions that exist between two organisms (i.e. flowering plant and bee, oxpecker and
rhino, and clownfish and sea anemone) (Figure 34). The alternative conceptions
associated with this item includes (i) both organisms benefit from the relationship equally
(AAAS Project 2061, n.d.), (ii) smaller organisms benefit more in mutualistic
relationships, and (iii) animals benefit more from mutualists relationships relative to
plants. For both 6th and 8th graders, the most commonly selected response was A which
is associated with the alternative conception that two organisms in a symbiotic
relationship experience the same benefits (see Table 30). The least selected response for both groups was choice D which is linked to the alternative conception that plants do not benefit as much as animals do within a mutualistic relationship. The 8th graders in this sample selected choice D at a higher frequency relative to their 6th grade counterparts which may represent a shift to an animal-centric alternative conception.

The following diagram includes examples of how organisms may interact with one another in a mutualistic relationship.

| Flowering plant and bee | Oxpecker and rhino | Clownfish and sea anemone |

Which statement best describes a mutualistic relationship?

a. Both organisms benefit from the relationship equally.

✔ b. Both organisms benefit from the relationship in various ways.

c. Smaller organism benefits more than larger organism.

d. Animals benefit more than plants.

*Figure 34. MS-LSCI item 20.*

*Table 30. Item 20 response pattern.*

<table>
<thead>
<tr>
<th>Response</th>
<th>6th Graders Selection Percentage</th>
<th>8th Graders Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>31.0</td>
<td>23.5</td>
</tr>
<tr>
<td>B</td>
<td>48.3</td>
<td>49.0</td>
</tr>
<tr>
<td>C</td>
<td>16.8</td>
<td>15.7</td>
</tr>
<tr>
<td>D</td>
<td>3.9</td>
<td>11.8</td>
</tr>
</tbody>
</table>
Growth and Reproduction

MS-LSCI item 25 is connected to growth, development, and reproduction of organisms in that the item asks student to think about how the body size of humans increases from birth to adulthood in the context of cellular division (AAAS Project 2061, n.d.) (see Figure 35). The alternative conceptions associated with this item are (i) cells grow in size but the number of cells remains constant, (ii) organisms grow in size when food accumulates in cells, and (iii) organism growth is linked to the accumulation of dead cells (AAAS Project 2061, n.d.). The response pattern for item 25 suggests that 6th graders were more drawn to the alternative conceptions that organisms grow when cells grow larger (response A) and dead cells take up space (response D) (see Table 31). When compared to the 6th graders, the 8th graders’ response pattern suggests that there may be a conceptual understanding shift related to cellular growth between the two grade levels. For example, 6th graders may be drawing on experiential knowledge about how hair and nails lengthen as dead cells accumulate to make a generalization about growth in organisms, whereas the 8th graders tended to select the response choice most closely aligned with cell theory.
As a human ages from birth to adulthood, the increase in body size is mostly the result of _________________.

- a. Cells growing larger.
- b. Existing cells creating new cells. **✓**
- c. New cells entering the body from food.
- d. Dead cells taking up space.

*Figure 35.* MS-LSCI item 25.

<table>
<thead>
<tr>
<th>Table 31. Item 25 response pattern.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
</tbody>
</table>

*Discussion*

The results of this study indicate that MS-LSCI performance was significantly lower for 6th graders relative to 8th graders and that six items can be linked to this difference in performance. These six items fell within the MS-LSCI’s core concepts of evolution and diversity, population interaction, and growth and reproduction. These items shed light on the contextual differences between 6th and 8th grade performance on MS-LSCI including differences in alternative conceptions held by each grade level.
Connecting Contextualized Item Difference and Response Patterns to Middle School Life Science Curriculum and Alternative Conceptions

The NGSS Middle Science Life Science concepts are organized by grade band and not individualized grades which allows grade level mapping to be determined at the discretion of the local education agencies. In states that have not adopted NGSS, an analogous grade banding structure is often present in state developed science standards. While there is local autonomy to control the middle school life science curriculum, similar patterns of concept sequencing are often found between different middle schools’ curriculum map at each grade level.

The middle school students in this study represented four schools located in one of two states; and of these states, one state follows state adapted NGSS, while the other state aligns instruction with state developed standards. Upon the examination of these schools’ life science curriculum maps, a general pattern of concept sequencing appeared regardless of which standards were utilized. Concepts related to the form and function of unicellular and multicellular organisms (i.e. MS-LSCI’s core concept of growth and reproduction) and cycles of matter and energy flow (i.e. MS-LSCI’s core concepts of energy and matter and population interactions) were explored in grades 6 and 7 followed by species and reproduction (i.e. MS-LSCI’s core concepts of growth and reproduction, inheritance, and evolution and diversity) in grade 8.
Evolution and diversity

In part, the arrangement of this curriculum may have contributed to the six items that provide the contextualized difference between 6th and 8th grade performance on the MS-LSCI. Since concepts of evolution and diversity are not introduced until 8th grade (NGSS Lead States, 2013), one would expect that 6th graders would tend to perform lower than 8th graders on the MS-LSCI items 3 and 4 which focus on natural selection and evidence of common ancestry.

For item 3, 6th graders had a higher tendency to select distractor choice associated with Lamarkism, whereas the 8th graders were more drawn to the distractor choice related to need-based changes in traits. While a shift from a Lamarkian perceptive to a needs-based perspective were displayed between the 6th and 8th graders, the results suggest that learners need additional support as they develop their scientific understanding of natural selection. Curriculum that targets how natural selection occurs at the species level over many generations, along with explicit instruction on how scientists utilize academic language to describe natural selection relative to a novice, may help middle school students shift to more expert-like conceptual understanding.

MS-LSCI item 4 examined students’ understanding of evidence of common ancestry specifically between chimpanzees and humans. This similar response pattern between 6th and 8th graders indicates that alternative conceptions related to common ancestry in the context of human and chimpanzees remain persistent between grades.
This persistence of alternative conceptions held about human-chimpanzee common ancestry has also been discovered at both the elementary (Buchan, Hejmadi, & Hurst, in press) and high school levels (Malone, Schuchardt, & Sabree, in review). This multi-grade band struggle to overcome these alternative conceptions calls for explicit and scaffolded curriculum spanning elementary to high school on human evolution and planned learning episodes that help students deconstruct the term ‘common ancestry’.

**Population interactions**

Concepts related to population interactions are either introduced in 6th grade or 7th grade depending on the school in this study. In this study’s sample, a larger proportion of the 6th graders attended schools that covered population interactions in 7th grade and not in 6th grade. This curriculum sequencing may have contributed to performance differences observed on the MS-LSCI items 15, 19, and 20 given that these items focus on resource competition, comparing carnivores and herbivores, and symbiotic relationships.

On items 15 and 20, both 6th and 8th graders in this sample displayed a higher propensity of selecting the distractor linked to the alternative conception that competition between organisms always involves interactions that are direct and aggressive and that two organisms in a symbiotic relationship experience the same benefits, respectively. These response patterns suggest that middle school curriculum should allow learners to explore the complexity of population interactions through activities that allow students to
examine diverse ecosystems rather than a few simplified models of ecosystems. This exploration may help students find evidence of organisms competing for resources in a nonaggressive/nondirect manner (i.e. indirect resource competition, interference competition, preemption, and allelopathy) and not equally benefiting from symbiotic relationships (i.e. service-resource relationships such as plant’s fruit being consumed by an animal and the seeds are dispersion in the animal’s excretions).

**Growth and reproduction**

While the sequencing of curriculum may be attributed to differences in item performance dealing with population interactions and evolution and diversity, the same argument cannot be made for MS-LSCI item 25 which focuses on growth and reproduction. Specifically, item 25’s content centers on how multicellular organisms’ growth is attributed to cellular division. Based on both states’ curriculum, 6th grade student would have studied the form and function of unicellular and multicellular organisms including cellular division and levels of organization in organisms. However, these 6th graders may be holding onto alternative conceptions that the 8th graders are not. In this sample, 58 percent of the 6th graders compared to only 36 percent of the 8th graders selected one of the following alternative conceptions related to the growth of multicellular organisms: (i) cells growth in size but the number of cells remains constant, (ii) organisms grow in size when food accumulates in cells, and (iii) organism growth is linked to the accumulation of dead cells. This shift in response pattern between 6th and
8th grade may suggest that a continued emphasis on cell theory as a unifying concept to life science curriculum aids middle school learners as they progress from experiential explanations to more conceptual understandings of growth in multicellular organisms (AAAS Project 2061; NGSS Lead States, 2013).

**Implications**

The MS-LSCI aims to measure a variety of fundamental life science core concepts that are introduced to students at various points in their time in middle school. This study’s results suggest that the MS-LSCI has the potential to distinguish middle school performance between grade levels, and therefore has the potential to be a valuable tool used by life science educators and researchers. The MS-LSCI can be used as a diagnostic assessment at different points throughout middle school which would allow student performance to serve as a focal point of lesson planning and differentiation. Middle school teacher teams can have targeted discussions about student progress before and after the instruction of the life science units allowing teachers the ability to identify areas of strength and growth. In the context of science education research, the MS-LSCI has the potential for being used in quasi-experimental studies. For example, the MS-LSCI can be employed to compare different instructional approaches by examining the difference between students’ pre- and post-test results. Research centered on grade-level comparisons of life science conceptual understanding may also benefit from the utilization of the MS-LSCI. The MS-LSCI has the potential to aid researchers
conducting longitudinal studies to track changes and shifts in middle schooler conceptual understanding, which in turn, can help identify curricular needs at each grade level.

**Future Research**

In order to further examine the MS-LSCI’s validity including generalizability, future research is needed to replicate and extend the finding in this study in a wide variety of middle school contexts and student samples. Therefore, additional field-testing of the 25-item MS-LSCI is needed, in part, to explore if similar patterns of performance are measured between middle school grade levels. In addition, of the six items linked to the contextualized differences between 6th and 8th graders, four items appear on the 25 item MS-LSCI. Further research is needed to examine if these four items are reliable at distinguishing middle school performance on the 25-item MS-LSCI. While this study excludes the analysis of 7th grade conceptual understanding changes and shifts due to the lack of significant MS-LSCI performance in comparison to 6th and 8th graders, additional research is needed to explore if the MS-LSCI can measure 7th grade level differences including alternative conceptions prevalence.

**Limitations**

While this study explores the possible curricular links between the contextualized differences and alternative conception response patterns between 6th and 8th grade performance on the MS-LSCI, further research is needed on how curricular mapping
influences performance on the MS-LSCI including the frequency of alternative conceptions held by students at different stages of instruction.

One area of improvement for this study centers on sampling. This study’s middle school student sample included students from only urban schools in the US Midwest. For future studies, the item field testing with middle school student should include students from diverse school settings including school situated in rural and suburban communities. By having a more diverse middle school sample, future studies would be positioned to discern if item performance is influenced by contextual differences between learning communities.
References


Chapter 5: Discussion of Three Studies

In this chapter, the connections between the three studies will be discussed. Additionally, the implications and limitations of the studies will be summarized, followed by recommendations for future research.

Overview of Studies

The goal in developing the MS-LSCI was to design a measurement instrument that centered on life science concepts that could provide measures related to the understanding of a large, diverse sample of middle-school aged students (grades 6-8). This concept inventory needed to be shown reliable and valid within the middle school population while also displaying the ability to discern the abilities of students who hold varying levels of life science conceptual knowledge. As a collective, the three studies provide evidence that progress towards attaining these overarching goals for the MS-LSCI have been made. Study one described how qualitative data from a multi-panel expert review and student interviews were used in the MS-LSCI’s item content validation and iterative refinement process. Study two described the MS-LSCI’s quantitative validation and item selection process through the utilization of Rasch modeling. When coupled, studies one and two describe the MS-LSCI’s mixed methods approach which is recommended during instrument development and validation (Liu, 2010; Schmeiser & Welch, 2006). Study three described the contextualized difference among 6th, 7th, and 8th
graders’ performance as measured by MS-LSCI; and thereby, provided evidence that this instrument has the capacity to discriminate between students who hold varying levels of life science conceptual knowledge.

Implications

The MS-LSCI, like other concept inventories, can help provide teachers and researchers with an understanding of each of their students’ naïve, alternative, existing, and intuitive conceptions (Bransford, Brown, & Cocking, 2000; Kuhn, Amsel, & O’Loughlin, 1988). Assessments that target students’ conceptual knowledge can help guide teachers’ instruction and curriculum decision making.

Pre-assessment

By using the MS-LSCI to investigate how students think about concepts prior to engaging in formal learning activities, an instructor can design lesson plans that support individual learners through the use of differentiation. That is, the MS-LSCI could be deployed at the beginning of the school year or the life science unit that is embedded within the course in order to measure and analyze where students’ initial conceptual understanding is so that the instructor can then develop an action plan to help scaffold learning (Marbach-Ad et al., 2010). In addition, middle school teacher teams can analyze the data collected from the MS-LSCI to make targeted curricular decisions for the inclusion of activities that target specific alternative conceptions of a particular cohort of students.
Learning activities that help students confront their alternative conceptions can be implemented after administration of the MS-LSCI. While some alternative conceptions are easier for students to confront (i.e. vernacular and factual misconceptions), other alternative conceptions are persistently engrained into students’ understanding of a phenomenon. For these persistent alternative conceptions, learning activities could be designed to cause perturbations and may include students using verbal, written, diagrammatic, and graphic explanations to represent their reasoning and supporting evidence.

The strategic use of the MS-LSCI can provide a pathway for two-way communication between the instructor and students. This communication provides students with opportunities to compare and receive feedback about their conceptual understanding with other students, teachers, and experts. In turn, two-way communication increases feedback opportunities, which are essential to support student conceptual understanding of science topics (Pellegrino, Chudowsky, & Glaser, 2001; Wilson & Scalise, 2006). Accordingly, learning activities should allow students to identify their alternative conceptions and help those students make accommodations to their reasoning that is aligned with scientific models of a phenomena.

*Pre/post assessment*

Concept inventories like the MS-LSCI can help instructors and researchers understand how new knowledge is acquired with respect to students’ prior knowledge.
through a pre-/post-test administration of the measurement instrument (Marbach-Ad et al., 2010; Zeineddin, & Abd-El-Khalick, 2008). Middle school teacher teams could have targeted discussions about student progress before and after the instruction of the life science units. These discussions would afford teachers the ability to identify areas of strength, in addition to detecting areas of growth, for the purposes of developing instructional goals and benchmarks to monitor growth. Moreover, this would allow teachers to develop learning activities that address common alternative conceptions throughout the unit from multiple student perspectives.

In the context of science education research, the MS-LSCI has the potential for being used in quasi-experimental studies. For example, the MS-LSCI could be employed to compare different instructional approaches by examining the difference between students’ pre- and post-test results. Research centered on grade-level comparisons of life science conceptual understanding may also benefit from the utilization of the MS-LSCI. The MS-LSCI has the potential to aid researchers conducting longitudinal studies that track changes and shifts in middle schooler conceptual understanding, which in turn could help identify curricular needs at each grade level.

Post-assessment

The MS-LSCI could serve as a summative assessment at the end of a course or end of the life science unit in order to measure performance patterns in the class as a whole, sub-populations within the class, and/or individual students (D'Avanzo, 2008).
Furthermore, because each of the distractor choices on a concept inventory are typically aligned with naïve, alternative, existing, and intuitive conceptions, an instructor could analyze the summative assessment results to find trends in how students think about specific concepts (Marbach-Ad et al., 2010). That is, if the results suggest little growth on a particular concept, then the instructor may need to restructure the learning activities for future courses and/or reteach the concept with those learners who may have struggled.

Additional Utility

While student interviews provide insight into students’ conceptual understanding, the time required to conduct interviews with every learner in a classroom is impractical for most educators. Concept inventories, including the MS-LSCI, offer a time efficient alternate to student interviews (Adam & Wieman, 2010; D’Avanzo, 2008; Garvin-Doxas, Klymkowsky & Elrod, 2007; Hestenes, Well & Swackhammer, 1992; Smith & Tanner, 2010).

As highlighted in studies one and two, assumptions related to the validity and reliably of the measurement of learning are often violated when instruments are used that have not been properly vetted using both qualitative and quantitative methods. To help circumvent the risk of making invalid inferences about conceptual understanding that influence their students’/subjects’ grades, academic standing, and access to instructional interventions, a high quality qualitative and quantitative review of all assessment item should include expert panel reviews, student interviews, and a psychometric evaluation.
However, conducting exhaustive qualitative and quantitative reviews for every assessment used in a classroom is time consuming and unrealistic. Concept inventories help mitigate this issue because they increase access to empirically developed and validated measurement instruments.

**Limitations**

The MS-LSCI, like all concept inventories, has inherent limitations. Due to the target population of this concept inventory (i.e. 6th, 7th, and 8th grade life science students), the MS-LSCI needed to be able to be administered during a typical class period. The MS-LSCI targets each of the NGSS life science disciplinary core ideas in middle school, however, the time constraints of a typical class period narrows the scope of concepts that can be covered. Additionally, the selected response format of the MS-LSCI, while chosen primarily for its ease of scoring, does not record a student’s conceptual understanding to the same degree that an instrument displaying open-ended or interview question formats would.

From a psychometric perspective, unidimensionality (i.e. a single conceptual construct) is critical for a valid measurement instrument. Due to its ability to measure a diverse set of life science core concepts across multiple grade levels, the MS-LSCI might appear multidimensional on the surface. However, all of the life science core concepts are underpinned by the unifying theory of evolution, and thus the MS-LSCI displays unidimensionality as verified by the Rasch model’s Item Fit statistics. In addition, the
NGSS standards for middle school life science are targeted to grade bands (i.e. 6th grade through 8th grade), rather than to a specific grade level. The core concepts of the MS-LSCI’s are aligned with those found in the NGSS standards, and as a result the MS-LSCI has broad applicability in that it can be used in a wide variety of middle school life science settings. Rather than measuring a narrow concept exhaustively, as is typical of many concept inventories, the MS-LSCI aims to measure a variety of fundamental life science core concepts that are introduced to students at various points during their middle school career.

A significant area of improvement for this study would be diversifying its sample. This study’s middle school student interviewee sample included students from one urban school in the US Midwest. For future studies, middle school student interviews should also include students from a variety of school settings, including schools situated in both rural and suburban communities. A more diverse middle school sample would position future studies to better discern if item performance is influenced by contextual differences between learning communities.

**Future Research**

Future research is needed to confirm that the findings in this study are generalizable and extend to a wide variety of middle school contexts and student samples. In addition, further research should examine whether or not the six items linked to the contextualized differences between 6th and 8th graders can reliably distinguish middle
school performance on the 25-item MS-LSCI. Due to the lack of significant MS-LSCI performance when compared to 6th and 8th graders, study three excluded the analysis of 7th grade conceptual understanding shifts. Additional research is needed to explore if the MS-LSCI is capable of measuring 7th grade level differences, including the prevalence of these students’ alternative conceptions.

Additionally, research is needed to determine how curricular mapping influences performance on the MS-LSCI, including the frequency of alternative conceptions held by students at different stages in instruction. The persistence of alternative conceptions between grade level and grade bands also demands further investigation. For example, there is a need to study how alternative conceptions related to human-chimpanzee common ancestry change between elementary, middle school and high school (Buchan, Hejmadi, & Hurst, in press; Ha & Cha, 2008; Malone, Schuchardt, & Sabree, in review; Stern & Hagay, 2005).

Additional work on the MS-LSCI is necessary prior to its utilization in classrooms. Final item formatting and item illustrations need to be completed. In developing a concept inventory, the final step is creating a user’s manual/guide (Liu, 2010). The following information is often included in concept inventory manuals: purpose and use of the inventory, table of test specifications, scoring guides, scoring category descriptions, and validity and reliability measures (Liu, 2010). Two versions of the user’s manuals need be created for the MS-LSCI: (i) a teacher’s manual, and (ii) a
researcher’s manual. These manuals should target each population’s needs, which will help contribute to the social validity of the MS-LSCI.
References


Complete References


178


