The Molecular Genetics Learning Progressions: Revisions and Refinements Based on Empirical Testing in Three 10th Grade Classrooms

Amber Nicole Todd
Wright State University

Follow this and additional works at: http://corescholar.libraries.wright.edu/etd_all
Part of the Biomedical Engineering and Bioengineering Commons

Repository Citation
THE MOLECULAR GENETICS LEARNING PROGRESSIONS: REVISIONS AND REFINEMENTS BASED ON EMPIRICAL TESTING IN THREE 10TH GRADE CLASSROOMS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By

AMBER NICOLE TODD
A.S., Cottey College, 2003
B.A., Mount Holyoke College, 2005

2013
Wright State University

Lisa Kenyon, Ed.D.
Dissertation Director

Mill Miller, Ph.D.
Director, Biomedical Sciences Ph.D. Program

R. William Ayres, Ph.D.
Interim Dean, Graduate School

Committee on Final Examination

Lisa Kenyon, Ed.D.

Steven Berberich, Ph.D.

Madhavi Kadakia, Ph.D.

James Tomlin, Ed.D.

Suzanne Franco, Ed.D.
Abstract


In the past few decades, there has been a large push for increasing scientific literacy (AAAS, 1989; AAAS, 1993; Achieve, 2013; NRC, 1996; NRC, 2012), especially in areas that are rapidly advancing, like molecular genetics. Much research has been done on student understandings of molecular genetics and the consensus is that the concepts are difficult both to learn and teach (Fisher, 1992; Horwitz, 1996; Kindfield, 1992; Lewis & Kattmann, 2004; Marbach-Ad & Stavy, 2000; Stewart et al., 2005; Venville & Treagust, 1998; etc.). Two learning progressions in molecular genetics have been produced (Duncan et al., 2009; Roseman et al. 2006), but both progressions are hypothetical as neither have been fully empirically tested. This study filled several gaps in molecular genetics research such as empirically testing the molecular genetics learning progressions in three 10th grade classroom contexts in different schools, determining the impact of curricular intervention units targeted to certain constructs of one of the progressions, and revising and refining the progressions based on empirical data.

The data collected show that 10th grade students fall on the extremely low levels of the progression prior to instruction and progress through the defined levels of the Duncan et al. (2009) progression for each construct. Students hold several lower and intermediate ideas that were added to the progression as new levels in each construct. It
was difficult to quantify the impact of the intervention units with quantitative data, but qualitative data suggest that certain ideas emphasized in the units such as a gene, protein, cell, trait scaffold and several specific examples of protein structures and functions were useful for students to understand ideas in molecular genetics. Additionally, two of the constructs of the Duncan et al. (2009) progression were divided into two new constructs each, and an entirely new construct was added to combine the Duncan et al. (2009) and Roseman et al. (2006) progressions. This is the first study to empirically test, revise, and refine all constructs of the Duncan et al. (2009) molecular genetics learning progression and to combine the Duncan and Roseman progressions into a single learning progression.
# Table of Contents

I. Introduction and Purpose .................................................................................................................1
   Scientific Literacy .........................................................................................................................1
   Literacy in Molecular Genetics ..................................................................................................4
   Learning Progressions .................................................................................................................11
   Revision and Refinement of Learning Progressions Based on
      Empirical Data .......................................................................................................................15
   The Molecular Genetics Learning Progressions .........................................................................18
   Purpose for Study .......................................................................................................................27

II. Methods .......................................................................................................................................32
   Study Context .............................................................................................................................32
   Intervention Units .......................................................................................................................36
   Data Collected ...........................................................................................................................39
   Theoretical Framework ...............................................................................................................42
   Development of Coding Schemes ...............................................................................................43
   Data Analysis .............................................................................................................................54
   Rubric to Learning Progression Revision and Refinement .........................................................64
   Statistical Analyses ....................................................................................................................65
   Reliability .....................................................................................................................................67
   Trustworthiness and Observer Effects .......................................................................................67
Empirically Testing and Revising the Molecular Genetics

Learning Progressions ................................................................. 73

Construct A ................................................................................. 74
Construct B .................................................................................. 80
Construct C .................................................................................. 86
Construct D .................................................................................. 96
Construct E .................................................................................. 102
Construct F .................................................................................. 108
Construct G .................................................................................. 114
  G1 ............................................................................................... 119
  G2 ............................................................................................... 124
Construct H .................................................................................. 127

Combining the Duncan et al. (2009) and Roseman et al. (2006) Progressions ......................................................... 133

Impact of Intervention Units .............................................................. 139

Protein Structure and Function ........................................................ 143
Gene, Protein, Cell, Trait Scaffold ........................................................................ 145

IV. Discussion ..................................................................................... 152

Empirically Testing and Revising the Molecular Genetics

Learning Progressions ...................................................................... 152
Impact of Intervention Units ................................................................. 159

V. Conclusion .......................................................................................... 170

VI. References .......................................................................................... 172

Appendix A ............................................................................................... 181

Appendix B ............................................................................................... 191
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Molecular genetics learning progression created by Roseman <em>et al.</em> (2006)</td>
<td>20</td>
</tr>
<tr>
<td>2. Averages of written assessments and interviews</td>
<td>77</td>
</tr>
<tr>
<td>3. Percent of students at each level of construct A</td>
<td>78</td>
</tr>
<tr>
<td>4. Percent of students at each level of construct B</td>
<td>84</td>
</tr>
<tr>
<td>5. Percent of students at each level of construct C</td>
<td>93</td>
</tr>
<tr>
<td>6. Percent of students at each level of construct D</td>
<td>99</td>
</tr>
<tr>
<td>7. Percent of students at each level of construct E</td>
<td>106</td>
</tr>
<tr>
<td>8. Percent of students at each level of construct F</td>
<td>112</td>
</tr>
<tr>
<td>9. Percent of students at each level of construct G1</td>
<td>120</td>
</tr>
<tr>
<td>10. Percent of students at each level of construct G2</td>
<td>125</td>
</tr>
<tr>
<td>11. Percent of students at each level of construct H</td>
<td>131</td>
</tr>
<tr>
<td>13. Average of interview responses for construct B</td>
<td>144</td>
</tr>
<tr>
<td>14. Average of interview responses for construct D</td>
<td>146</td>
</tr>
<tr>
<td>15. Average of interview responses for construct G1</td>
<td>148</td>
</tr>
<tr>
<td>16. Average of interview responses for construct G2</td>
<td>150</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Molecular Genetics Learning Progression Created by Duncan <em>et al.</em> (2009)</td>
<td>22</td>
</tr>
<tr>
<td>2. Classroom Contexts Used for Study</td>
<td>33</td>
</tr>
<tr>
<td>3. Codes Developed for Constructs</td>
<td></td>
</tr>
<tr>
<td>3.1 Codes Developed for Construct A</td>
<td>45</td>
</tr>
<tr>
<td>3.2 Codes Developed for Construct B</td>
<td>46</td>
</tr>
<tr>
<td>3.3 Codes Developed for Construct C</td>
<td>48</td>
</tr>
<tr>
<td>3.4 Codes Developed for Construct D</td>
<td>50</td>
</tr>
<tr>
<td>3.5 Codes Developed for Construct E</td>
<td>52</td>
</tr>
<tr>
<td>3.6 Codes Developed for Construct F</td>
<td>55</td>
</tr>
<tr>
<td>3.7 Codes Developed for Construct G1</td>
<td>57</td>
</tr>
<tr>
<td>3.8 Codes Developed for Construct G2</td>
<td>59</td>
</tr>
<tr>
<td>3.9 Codes Developed for Construct H</td>
<td>61</td>
</tr>
<tr>
<td>4. Empirical Revisions and Refinements of Constructs</td>
<td></td>
</tr>
<tr>
<td>4.1 Empirical Revisions and Refinements of Construct A</td>
<td>75</td>
</tr>
<tr>
<td>4.2 Empirical Revisions and Refinements of Construct B</td>
<td>81</td>
</tr>
<tr>
<td>4.3 Empirical Revisions and Refinements of Construct C1</td>
<td>88</td>
</tr>
<tr>
<td>4.4 Empirical Revisions and Refinements of Construct C2</td>
<td>90</td>
</tr>
<tr>
<td>4.5 Empirical Revisions and Refinements of Construct D</td>
<td>97</td>
</tr>
<tr>
<td>4.6 Empirical Revisions and Refinements of Construct E</td>
<td>104</td>
</tr>
</tbody>
</table>
4.7 Empirical Revisions and Refinements of Construct F ........................................109
4.8 Empirical Revisions and Refinements of Construct G1.......................................115
4.9 Empirical Revisions and Refinements of Construct G2.......................................117
4.10 Empirical Revisions and Refinements of Construct H ......................................128

5. Hypothetical Construct I ........................................................................................138
Acknowledgments

I would like to thank my advisor, Dr. Lisa Kenyon for allowing me to work in her lab, for encouraging me to pursue my own interests and develop my own project, as well as work with the MoDeLS (Modeling Designs for Learning Science) and Practices teams. I wish to thank her for her mentoring, support, encouragement to attend conferences, and conversations during my time in her lab. I also want to thank Dr. Steven Berberich for allowing me to work in his lab completing my non-dissertation research and for all of his extremely helpful guidance and support through this entire program. I wish to thank my committee members Dr. Madhavi Kadakia, Dr. James Tomlin, and Dr. Suzanne Franco for all of their questions, feedback, guidance, help, and time. I want to additionally thank the Biomedical Science Ph.D. Program, Dr. Gerald Alter, Dr. Mill Miller, Karen Luchin, and Diane Ponder for all of their assistance and willingness to allow me to complete my dissertation research in science education.

I would like to thank the teachers, students, administrators, and staff that I worked with collecting data. Going in to the classrooms collecting data and listening to student ideas and hearing great teacher-student interactions was such a joy and a constant reminder of how educational research can impact lives. A special thank you to Dr. Melissa Schen for help with statistics as well as the other graduate students and undergraduates in the Kenyon lab for help with interviews, data collection, and interpretations, especially Brooke Andrews Ratliff, Mackenzie English, and Jeannette Loyer.
Last, but certainly not least, I especially wish to thank my family for their constant love and support. Jason, Amelia, Mom, Dad, Shawn (& Sally, Kristi, Lucas, Alyssa), Micah, Lonnie, Sue (& Matt, Molly, Meagan, Mike, Amanda, Blaine) - thank you all so much! I would not have been able to do this without every single one of you. I love you all and am very grateful to have you all in my life. Especially my amazing husband and adorable daughter!
I. Introduction and Purpose

Scientific Literacy

The term “scientific literacy” has its roots in the 1950’s era of the space race with the Soviet Union and Sputnik (Laugksch, 2000). During this time, Americans were becoming concerned that their children were not receiving an adequate education to be able to compete with foreign science and technology powers, such as the Soviet Union (Hurd, 1958). Increasing scientific literacy was seen as a way to combat the potential problem (Hurd, 1958; Waterman, 1960). Over the next few decades, much was written about scientific literacy both in terms of the concept itself and what was actually meant by the concept (reviewed in Laugksch, 2000).

Multiple meanings of scientific literacy have emerged over the years and the term still remains an “ill-defined and diffuse concept” (Champagne & Lovitts, 1989). Despite the fact that this seemingly simple concept has so many facets and interpretations, scientific literacy generally “stands for what the general public ought to know about science” (Durant, 1993, p. 129), and “commonly implies an appreciation of the nature, aims, and general limitations of science, coupled with some understanding of the more important scientific ideas” (Jenkins, 1994, p. 5345).

The popularity of scientific literacy has waxed and waned over the years (reviewed in Laugksch, 2000), but the concept was again placed in the recent spotlight with the establishment of Project 2061 by the American Association for the Advancement of Science (AAAS). Project 2061 aims at reforming science, mathematics, and
technology education in the United States to increase scientific literacy (American Association for the Advancement of Science [AAAS], 1989). Indeed, the product of Phase II (of three phases of the project) is titled *Benchmarks for Science Literacy* (AAAS, 1993). The Phase II publication included the goals for achieving scientific literacy by grade levels, indicating what benchmarks students should achieve by certain grades in order to achieve scientific literacy by the time they graduate from high school (AAAS, 1993).

Additionally, the National Research Council (NRC) very recently released *A Framework for K-12 Science Education* (2012) which promoted a new approach to K-12 science education to help increase scientific literacy. The NRC approach included concurrently teaching crosscutting concepts, scientific and engineering practices, and the disciplinary core ideas all through K-12 science. The authors outlined a broad set of expectations for students in grades K-12; the purpose was that they would be used to inform the development of new K-12 science standards, science curriculum science instruction methods, assessments, and professional development. Indeed, the expectations outlined in the *Framework* have already been used to develop the Next Generation Science Standards (Achieve, 2013). Researchers, teachers, and curriculum developers are currently reviewing the *Framework* and the Next Generation Science Standards to inform changes to curriculum, assessments, instruction methods, and professional development. These very recent changes address the issue of increasing scientific literacy through K-12 science education in schools.

There are several reasons to promote scientific literacy. The reasons can be broken down into two broad categories or views: macro and micro. The macro view is
concerned with how scientific literacy is important for the economy of a nation; it was the predominant view during the establishment of the term “scientific literacy” during the space race era and also appears in *A Framework for K-12 Science Education* (2012). The macro view is based on the belief that scientists, engineers, and technically trained people develop and sustain the technology of a nation, and the best way to obtain a steady supply of people for these occupations is through the production of a scientifically literate population. The view also includes the argument that increased scientific literacy of a population will increase the public support of science itself, decrease unrealistic expectations of science, and lead to better policy-making decisions in science when the public casts ballots regarding scientific decisions (Laugksch, 2000).

The micro view is centered around how scientific literacy is important to individuals themselves. In the current science and technology-dominated society, it is extremely advantageous for an individual to be scientifically literate. Science and technology influence such daily decisions as diet, smoking, and vaccinations. Having a clear grasp of science will also help individuals identify the difference between true science and pseudo-science concepts which infiltrate current society (Laugksch, 2000). The Royal Society (1985, p. 10) documented that “an uninformed public is very vulnerable to misleading ideas on, for example, diet or alternative medicine.” As the use of technology and scientific and technological advances continue to increase, scientific literacy is becoming more and more necessary for individuals to be able to confidently and competently deal with advanced biological topics as they arise in his or her daily life (Laugksch, 2000). It is clear that increasing scientific literacy is important, regardless of which view reason is the most compelling. As such, both of these categories are
mentioned in *Science for All Americans*, the product of Phase I of Project 2061 (AAAS, 1989).

**Literacy in Molecular Genetics**

Helping students become scientifically literate is certainly a challenge (AAAS, 1989; AAAS, 1993; National Research Council [NRC], 1996; NRC, 2012), but an even bigger challenge is helping students become scientifically literate in areas that are rapidly advancing. Molecular genetics is one such rapidly advancing area due to the recent scientific and technological advances like the sequencing of the human genome, genetic screening, genetically modified organisms, and stem cell research, among others. Molecular genetics is a complex topic and scientists are continually contributing to the wealth of information already obtained.

Much research has been done on student understandings of molecular genetics and the consensus is that the concepts are difficult both to learn and teach (Fisher, 1992; Friedrichsen & Stone, 2004; Horwitz, 1996; Kindfield, 1992; Lewis & Kattmann, 2004; Lewis & Wood-Robinson, 2000; Marbach-Ad & Stavy, 2000; Stewart, Cartier, & Passmore, 2005; Stewart & Van Kirk, 1990; Venville & Treagust, 1998; Wynne, Stewart, & Passmore, 2001). Literacy in molecular genetics is especially important because the general public is beginning to encounter molecular genetics during the course of their everyday lives. In this new century, molecular genetics will likely have the most immediate and direct impact on an individual of any science area. A lack of understanding of molecular genetics translates to being unable to properly understand and benefit from new technologies such as genetic screening (Gollust, Wilfond, & Hull, 2003; Hull & Prasad, 2001; Lewis & Wood-Robinson, 2000). In addition, the public has been
called upon to make informed decisions about such topics as cloning, gene therapy, and stem cell research. Research repeatedly documents that the general public is making uninformed decisions based on their lack of literacy in molecular genetics (Fisher, 1992; Garton, 1992; Kindfield, 1992) and that many high school graduates who have passed required life science courses are even ill-equipped to make informed decisions about topics in molecular genetics (Lanie et al., 2004; Lewis & Wood-Robinson, 2000).

In 2005, Stewart, Cartier, and Passmore described that molecular genetics literacy included understanding and integrating three conceptual models in genetics. The models are the genetic (sometimes called inheritance, Mendelian, classical, or transmission genetics) model, meiotic model, and the bio-molecular model (hereafter referred to as the molecular model). The genetic model explains the patterns of correlation between genes and traits. The meiotic model explains the cellular processes by which genetic information is transferred between parents and offspring. The molecular model explains the mechanisms inside the cell by which genes give observable traits or physical effects. Stewart et al. (2005) explained that literacy in molecular genetics consists of understanding each of the three models and being able to integrate them into coherent explanations of genetic phenomena. Given the complexity of not only understanding the separate models but also integrating them, it is not surprising that several studies have shown that students have problems with the tasks (Cartier, 2000; Freidenreich, Duncan, & Shea, 2011; Kindfield, 1994; Wynne, et al., 2001).

The lack of molecular genetics literacy in students can be attributed to two main factors: complexity of content and current classroom instruction. The content is considered complex due to the hierarchical levels of organization (genes, proteins, cells,
tissues, organs, etc.) where one level forms the next level of organization (Duncan & Reiser, 2007; Hmelo-Silver & Azevedo, 2006; Horwitz, 1996; Knippels, 2002; Simon, 1996). Also, interactions at the micro-level (protein-protein, cell-cell) give rise to the macro-level patterns that one can actually see, such as physical traits (Casti, 1994; Horwitz, 1996). To further complicate molecular genetics content, there is an informational level (genes) and a biophysical layer (proteins, cells, tissues, etc.) that are hierarchical (Simon, 1996), which has been termed “hybrid hierarchical” (Duncan & Reiser, 2007). Additionally, current classroom instruction in molecular genetics consists mainly of memorizing processes and vocabulary terms instead of emphasizing the big ideas and understanding the mechanisms behind them (AAAS, 2005; Duncan & Reiser, 2007; Kurth & Roseman, 2001; NRC, 2012; Venville & Treagust, 1998).

There has been significant research on student conceptions in molecular genetics under normal classroom instruction. Students struggle with the levels of organization and generally fail to understand that genes do not directly code for observable traits. They do not grasp that genes simply code for a sequence of amino acids for a protein (Duncan & Reiser, 2007; Marbach-Ad, 2001; Venville, Gribble, & Donovan, 2005; Venville & Treagust, 1998). They also do not understand how observable traits come from microscopic interactions at the lower levels of organization. A study by Marbach-Ad and Stavy (2000) documented that students were not able to explain at the molecular and cellular level how visible traits come about. Students tended to connect genotype and phenotype by explaining that the genotype “gives” the phenotype; that is, the genotype directly determines the phenotype, completely bypassing the role of proteins (Duncan & Reiser, 2007; Lewis & Kattmann, 2004). Bypassing proteins in the process is not
surprising given how little students know about proteins and their role in molecular genetics (Duncan, 2007; Rogat & Krajcik, 2006).

Because of student misconceptions of molecular genetics under normal classroom instruction, Dougherty (2009) proposed an “inverted” curriculum where more complex topics like polygenic traits are taught first and more simple models of inheritance patterns are discussed later. Dougherty (2009) explained that the complex traits are the predominant traits in humans and other organisms, yet are rarely discussed in curriculum. He contended that by learning about the simpler models first, students hold on to a more simplistic view of inheritance patterns (such as simple Mendelian dominant/recessive patterns) and fail to understand the more complex and biologically relevant polygenic traits and how the environment impacts genetics. Although the inverted curriculum may help students gain experience with more complex inheritance patterns which are often left out of the current curriculum, the suggestion contradicts several studies which explained that students struggle to understand and explain even basic inheritance patterns in the genetic model (Cartier & Stewart, 2000; Freidenreich et al., 2011; Lewis, Leach & Wood-Robinson, 2000; Tsui & Treagust, 2007; Wynne et al., 2001). If students struggle with understanding and explaining basic patterns, it may be extremely difficult to get students to first adequately understand complex patterns, especially at younger grades. Freidenreich et al. (2011) explained that although discussing complex patterns of inheritance is important and should be added to current curriculum, the authors’ work and the work of others suggests that the more complex patterns should be introduced in high school and not in middle school as Dougherty (2009) suggested.
Duncan (2007) also developed a cognitive model that outlined the types of knowledge that are critical for reasoning in molecular genetics. This model is based on data collected while helping undergraduate college students with the myriad of problems with molecular genetics education. Duncan (2007) described two types of domain-specific knowledge: heuristics and explanatory schemas. The heuristics included important concepts and relationships in molecular genetics while the schemas included important mechanisms in molecular genetics. Three heuristics were found to be important: genes-code-for-proteins, proteins-as-central and effects-through-interaction. The heuristics can be used across the field of molecular genetics to reason that genes code for proteins (and not traits), that proteins are key intermediate step between genes and traits, and that protein effects are mediated through interactions with other proteins. Duncan (2007) also described nine explanatory schemas: inhibit, activate, translation, regulation-of-gene-expression, catalyze, transport, receptor, structural, and structure-function. The explanatory schemas are key mechanisms for students to understand across the field of molecular genetics and all are related to the role proteins play in cells.

With Stewart et al.’s (2005) description of molecular genetics literacy consisting of the three inter-related conceptual models, the information about student misconceptions under normal classroom instruction, and Duncan’s (2007) development of a cognitive model for molecular genetics reasoning, numerous studies target increasing molecular genetics literacy in students by implementing various classroom interventions. The following is a brief summary of the more successful classroom interventions:
• The addition of bead and/or illustration models to an 11\textsuperscript{th} and 12\textsuperscript{th} grade curriculum was found by Rotbain, Marbach-Ad, & Stavy (2006) to increase knowledge in molecular genetics.

• Twelfth grade students gained a deeper understanding of genetics due to the addition of a web-based bioinformatics intervention (Gelbart & Yarden, 2006).

• Tsui & Treagust (2007) found that multiple representations of concepts in genetics did increase 10\textsuperscript{th} and 12\textsuperscript{th} grade students’ understandings of concepts, however only four of the nine students interviewed constructed ideas that were intelligible, plausible, and fruitful after the classroom intervention.

• Elkund, Rogat, Alozie, & Krajcik (2007) found modest gains in 9\textsuperscript{th}/10\textsuperscript{th} graders’ understandings of molecular genetics after implementation of an intervention unit.

• Interestingly, Venville & Donovan (2007) implemented an intervention in a 2\textsuperscript{nd} grade classroom that introduced DNA and genes and found that students that young could develop understandings of inheritance and concepts of DNA and genes.

• Addition of animations and illustrations into an 11\textsuperscript{th} and 12\textsuperscript{th} grade curriculum was found to increase general student knowledge in molecular genetics, with animations increasing knowledge more than the illustrations (Marbach-Ad, Rotbain, & Stavy, 2008; Rotbain, Marbach-Ad, & Stavy, 2008).

• A computer-based intervention called BioLogica also produced significant learning gains in high school students in genetics (Horwitz, Gobert, Buckley, & O’Dwyer, 2010).
• Duncan, Freidenreich, Chinn, & Bausch (2011) found that 7th grade students can generate the *genes-code-for-proteins* and *proteins-as-central* heuristics and use them to reason about molecular genetics; however development of this knowledge was highly dependent on the quality of classroom instruction.

• Freidenreich *et al.* (2011) found that an eight week intervention unit helped 6th-8th grade students increase their understanding of each of the three models in genetics.

• The addition of an intervention unit focusing on the core mechanisms and the important role proteins play in molecular genetics was found to increase 9th grade students’ understandings of molecular genetics (Duncan, 2006; Duncan & Tseng, 2011).

It is clear that instructional interventions are making some progress towards increasing molecular genetics literacy in students while normal classroom instruction is struggling to meet the needs of both students and teachers in learning and teaching genetics. The most successful interventions described above contained focused instruction on the core mechanisms behind molecular genetics, greater emphasis on the role of proteins in cells, animations and illustrations, and models. All interventions are consistent with the recommended changes to the current science curriculum by the reform-based science education movement (AAAS, 1989; AAAS, 1993; NRC, 1996; NRC, 2012). However, learning progressions are a key component of success in the reform of science education (Board on Science Education, 2010; NRC, 2005; NRC, 2012). As such, two learning progressions in molecular genetics have been recently published (Duncan, Rogat, & Yarden, 2009; Roseman, Caldwell, Gogos, & Kurth, 2006).
Learning Progressions

The general reform in science education movement, establishment of Project 2061 and *A Framework for K-12 Science Education*, has not only called for increasing scientific literacy (AAAS, 1989; NRC, 2012), but also called for better alignment among curriculum, instruction, and assessment in the classroom (NRC, 2005, 2007). The NRC reports *Taking Science to School: Learning and Teaching Science in Grades K-8* (2007) and *A Framework for K-12 Science Education* (2012), among others, posited that learning progressions play a key role in the curriculum, instruction, and assessment reform (Board on Science Education, 2010; NRC, 2005). Indeed, authors of the Framework explicitly noted that the “core ideas and their related learning progressions are key organizing principles for the design of the framework” (NRC, 2012, p. 26) and, thus, the Next Generation Science Standards (Achieve, 2013).

Learning progressions are a current “hot topic” in science education but are not a new idea. They generally describe “successively more sophisticated ways of reasoning within a content domain that follow one another as students learn” (Smith, Wiser, Anderson, & Krajcik, 2006, p. 1) and may describe content, practices, or both in a single progression (Corcoran, Mosher, & Rogat, 2009; NRC, 2007; Smith *et al.*, 2006). The learning progressions are similar to findings from other studies that have examined the development of children’s ideas over time (Brown & Campione, 1994; Bruner, 1960; Carpenter & Lehrer, 1999), but differ in that the learning progressions contain several distinct characteristics (reviewed in Duncan & Hmelo-Silver, 2009).

Four important theoretical and structural characteristics of science learning progressions were identified in a panel discussion on science learning progressions
organized by the Center on Continuous Instructional Improvement, the Consortium for Policy Research in Education in 2008 (Corcoran et al., 2009) and the previously mentioned NRC Taking Science to School report (2007) authors. The learning progression characteristics include the following 1) they are only focused on a few content ideas and practices, but may have the practices combined, 2) they contain upper and lower bounds which describe what the students should be able to attain at the end of the progression and what knowledge the students have when they enter the progression, 3) they identify varying levels of achievement in terms of learning performances between the two bounds, 4) the achievement described is reached through targeted instruction and curriculum (reviewed in Duncan & Hmelo-Silver, 2009).

The first two characteristics of learning progressions are fairly self-explanatory, but the third and fourth characteristics deserve some additional explanation. Learning progressions identify varying levels of achievement in terms of learning performances, which are grounded in research on how students actually understand specific scientific ideas. Because the ideas are based on how students learn the content, the intermediate steps in a progression may vary from canonical knowledge of the subject or even be scientifically inaccurate. The progression steps instead focus on deepening understandings and increasing complexity of ideas over time and can be seen as productive stepping-stones that position students in a better place to reach increasingly more complex ideas. Because learning progressions may include these scientifically inaccurate (yet productive) understandings as intermediate steps and because progressions are being used to inform design on standards, curriculum, instruction, and assessments, it is very controversial how much, if any, of the inaccurate (yet productive)
intermediate ideas to include in standards, curriculum, and instruction (Corcoran et al., 2009; Duncan & Rivet, 2013; Shea & Duncan, 2013; Wiser, Smith, Doubler & Asbell-Clarke, 2009).

Learning progressions describe varying levels of achievement as previously described, but the achievement must be reached through targeted instruction and curriculum. That is to say, students do not naturally attain the achievement levels without the proper instructional and curricular support; student attainment of the levels is not assured, even with proper support. It is also important to remember that although learning progressions appear to be linear, in progressing from level to level and increasing in sophistication, they are not necessarily linear. Students may take one of several different paths to increase their sophistication and need not necessarily follow the single linear path described in the progression. A single student’s progress “is likely more akin to ecological succession than to constrained lock-step developmental stages” (Duncan & Hmelo-Silver, 2009).

With the push for science education reform and the suggestion that learning progressions may help, several learning progressions in a variety of different fields, including molecular genetics, have been published (Alonzo & Steedle, 2009; Berland & McNeill, 2010; Catley, Lehrer, & Reiser, 2005; Duncan et al., 2009; Lee & Liu, 2010; Lehrer & Schauble, 2000; Mohan, Chen, & Anderson, 2009; Plummer & Krajcik, 2010; Roseman, et al., 2006; Schwarz et al., 2009; Smith et al., 2006; Songer, Kelcey, & Gotwals, 2009; Stevens, Delgado, Shin, & Krajcik, 2007; Stevens, Delgado, & Krajcik, 2010). The production of a learning progression in a specific area involves synthesizing the research on student learning in that area as well as conducting empirical studies of the
progression itself. The empirical studies can come before, during, or after the creation of a progression but must be done at some point to validate the progression. It is important to note that although a progression takes all the existing research into consideration, it remains a hypothetical model of learning until empirically validated. Empirical studies of the progression then lead to multiple iterative rounds of revisions and refinement of the progression based on the classroom data obtained. Completed learning progressions also contain instructional materials that have been shown to support the progression of students through the levels of the progression as well as assessments (Duncan & Hmelo-Silver, 2009).

Despite the plethora of recent research on learning progressions, several unresolved issues remain. Progressions can differ greatly in terms of grain size. Some progressions focus only on a few years of instruction (Schwarz et al., 2009; Songer et al., 2009) while others contain a wide range of grades (Duncan et al., 2009; Mohan et al., 2009). The number of levels included and the speed at which these levels are achieved often differ as well. Also differing is the grain size of the content idea or practice itself. Another issue involves how integral the curriculum and instruction are to the progression. As previously mentioned, students need targeted curriculum and instruction to be able to move through the levels of the progression; however, the curriculum and instruction related to progressions can range from extensive interventions (Schwarz et al., 2009; Songer et al., 2009) to no interventions beyond normal classroom instruction (Duncan et al., 2009; Mohan et al., 2009). A final issue involves how best to validate the learning progressions given different classroom settings, instruction, and individual student history, all of which play a role in student achievement. However, it is unclear how to
take the school and student level variables into account when validating a progression and evaluating student performance related to a learning progression (reviewed in Duncan & Hmelo-Silver, 2009). Regardless of the unresolved issues, it is clear that learning progressions are a valuable tool in science education and will be useful to inform the field about how to better align standards, curriculum, instruction, and assessment.

**Revision and Refinement of Learning Progressions Based on Empirical Data**

As previously described, a learning progression remains a hypothetical model of learning until empirically validated. Empirical studies can come before, during, or after learning progression construction but must be done at some point to validate the progression. It is also worth noting that empirically validating a progression is not a one-time study; revisions and refinements of progressions happen through multiple iterative rounds of empirical studies. Though the methods used to empirically validate a progression may vary depending on the stage of completion of the progression, the fundamental bond that all learning progression revisions and refinements share is the need to correlate empirical data with hypothetical models (Shea & Duncan, 2013).

Though learning progressions are in their infancy, researchers are starting to revise and refine, and thus empirically validate a few. The methods for refinement from empirical data, however, are not always documented. Mohan *et al.* (2009) simply noted that “the researchers reflected on how best to revise the levels to capture the types of responses to those items.” Stevens *et al.* (2010) provided a side-by-side comparison of their initial hypothetical progression and empirically refined progression and explained the differences in detail, but did not explain how they decided upon the specific revisions and refinements. Alonzo and Steedle (2009) described how they used empirical data to
add and condense levels of their specific progression but fell short of providing general heuristics for the revision and refinement of progressions based on empirical data.

Shea & Duncan (2013) recently provided both detailed examples of revisions and refinements to the Duncan et al. (2009) molecular genetics learning progression based on empirical data as well as general heuristics that researchers may use across disciplines to revise and refine progressions based on their own empirical data. The heuristics described how to use empirical data to both revise and refine the levels of a learning progression (adding, removing, splitting, combining levels) and to characterize relationships between multiple constructs within a single progression. The authors contended that levels should be added when “the new ideas are directly related to the construct, represent an important conceptual shift, and/or afford instructional leverage.” They also cautioned that intermediate levels should be productive stepping-stones that position students in a better place to reach increasingly more complex ideas instead of a massive list of incremental understandings (Shea & Duncan, 2013). Productive stepping-stones in learning progressions are valuable to teachers because they provide pedagogical content knowledge (PCK). Shulman (1987, pp. 15) described the distinguishing characteristic between expert and novice teachers is the experts’ “capacity of a teacher to transform the content knowledge he or she possesses into forms that are pedagogically powerful and yet adaptive to the variations in ability and background presented by students.” The productive stepping-stones in learning progressions are transformations in content knowledge to pedagogically powerful and adaptive forms that students of various abilities do articulate. Having knowledge and understanding of these stepping-stones in content areas can help teachers prepare helpful and productive instruction for students.
Levels should be removed if empirical data show no (or very few) students at that level. The authors warned that removal of upper levels should be partially informed by expectations of what is reasonable for students to be able to master and that a lack of students at the upper level may indicate the need for instructional materials that target the corresponding specific idea of the progression. Another concern was that a lack of students at an intermediate level may indicate that students quickly move through an idea and that the idea is difficult to capture in small sample sizes (Shea & Duncan, 2013). Although Shea & Duncan (2013) did not explicitly mention distinct heuristics regarding splitting levels, they indicated that the same heuristics for adding levels (idea directly related to the content, representing an important conceptual shift, or affording instructional leverage) apply to splitting a level into two (or more) distinct levels. The authors then provided the heuristics for combining levels: the levels do not discriminate between abilities and that combining the levels would not result in a loss of information. Item difficulties in a Wright Map (Wilson, 2005) can be used to identify levels that have a similar ability level. It may be advisable to combine levels that have the same ability level, but an informed judgement must be made to determine if combining levels would result in a loss of valuable information. Sometimes ideas that have a similar ability level represent conceptually different ideas and combining these levels would result in a loss of valuable information (Shea & Duncan, 2013).

Heuristics to characterize relationships between multiple constructs within a single progression are a bit more complicated. If there are multiple constructs within a progression, the constructs are likely related and it is extremely likely that one (or more) construct(s) influences other constructs (Plummer & Krajcik, 2010; Schwarz et al., 2009;
Shea & Duncan, 2013; Wilson, 2009). Shea & Duncan (2013) described five categories of contingencies between two constructs: not connected, weakly connected, connected, strongly connected, and intertwined. They determined descriptions for each of these categories as they related to two of eight constructs and categorized students accordingly. Using the categories of contingencies, the authors were able to reason the extent of the relationship between the two constructs. Although the authors did not explicitly provide heuristics regarding identifying relationships between constructs, they stated “we believe that a similar approach could be used with larger numbers of constructs at play” as well as acknowledged the challenge of identifying relationships with multiple constructs (Shea & Duncan, 2013). Having heuristics to guide data-driven revisions and refinements to learning progressions is very helpful considering the emphasis that has been placed on progressions for aligning standards, curriculum, instruction, and assessment.

**The Molecular Genetics Learning Progressions**

Two learning progressions in the molecular genetics content area have been produced in response to the current reform-based movement in science education (Duncan *et al.*, 2009; Roseman *et al.*, 2006). Both progressions span grades 5-10 and focus on the molecular genetics content area, each including the three inter-related conceptual models of genetics (genetic, meiotic, and molecular). While neither progression explicitly includes any scientific practices, both implicitly include mechanistic models. The learning progressions are also both hypothetical as neither have been empirically tested across all grades included (reviewed in Duncan, 2011). The middle school expectations in the Duncan *et al.* (2009) learning progression have been recently tested and reported in a paper by Freidenreich *et al.* (2011). The longitudinal
study was designed to test the middle school and high school levels of the Duncan et al. (2009) progression; however the paper only included the first part of the findings with middle school students (grades 6-8) and in only one classroom context. The authors used their findings with middle school students to revise and refine a portion of their progression based on the empirical data (Shea & Duncan, 2013). The revisions will be discussed in detail later. No portions of the Roseman et al. (2006) progression have been empirically tested.

Roseman et al. (2006) developed the first learning progression in molecular genetics (Figure 1). The progression focuses on the two main functions of DNA: determining an organism’s characteristics (ideas denoted by a blue triangle pointing to the bottom right in Figure 1), and transferring genetic information between parents and offspring (ideas denoted by a red triangle pointing to the top left in Figure 1). The authors argued that molecular genetics instruction should deviate from normal classroom instruction in two ways: introducing proteins before DNA, and introducing the molecular model before the meiotic or inheritance models. Normal classroom instruction typically introduces proteins after discussion of DNA structure, replication, transcription, and translation. Roseman et al. (2006) argued that understanding the role proteins play in producing observable traits (the intermediate between genes and traits) is so important (Duncan, 2007; Duncan & Reiser, 2007; Lewis & Kattmann, 2004) that students must understand proteins and their functions in the cell before describing the function of DNA and how it gives observable traits. Additionally, they argued that first understanding the structure and function of DNA and proteins (the molecular model) will help students better understand the roles of genes, chromosomes, and alleles (the meiotic and genetic
The information passed from parents to offspring is coded in DNA molecules. DNA molecules are long chains linking just four kinds of smaller molecules, whose precise sequence encodes genetic information.

Genes are segments of DNA molecules. Each DNA molecule contains thousands of discrete genes.

The genetic information stored in DNA is used to direct the synthesis of the thousands of proteins that each cell requires.

A change in a single atom in the DNA molecule... can... change the protein that is produced.

Insertions, deletions, or substitutions in DNA can alter genes. Amutation in a DNA segment may not make much difference in the operation of the cell, may fatally disrupt it, or may change it in a significant way.

When mutations occur in sex cells, they can be passed on to all cells in the resulting offspring; if mutations occur in other cells, they can be passed on to descendant cells only.

Heritable characteristics ultimately produced in the development of an organism can be observed at molecular and whole-organism levels—in structure, chemistry, or behavior.

All matter is made up of atoms... Atomss may stick together in well-defined molecules or may be packed together in large arrays. Different arrangements of atoms into groups compose all substances.

The work of the cell is carried out by the many different types of molecules it assembles, mostly proteins.

Cells repeatedly divide to make more cells for growth and repair.

Some faulty operations of body processes are known to be caused by altered genes. They may have a direct, obvious effect, such as causing easy bleeding, or they may only increase the body's susceptibility to developing particular diseases, such as clogged arteries or mental depression.

Changes in DNA (mutations) occur... When mutations occur in sex cells, they can be passed on to all cells in the resulting offspring; if mutations occur in other cells, they can be passed on to descendant cells only.

models) because the roles are more abstract. Under normal classroom instruction, students are typically taught the molecular model last.

Duncan et al. released an additional learning progression in molecular genetics in 2009 (Table 1). The Duncan et al. (2009) progression has some similarities to and differences from the Roseman et al. (2006) progression. The Duncan et al. (2009) progression documented that proteins and their functions should be introduced before discussing DNA so that students understand how the products of genes (proteins) do the work of the cell to bring about observable traits or physical effects. However, unlike the previous learning progression, Duncan et al. (2009) argued that the three conceptual models of genetics should be taught concurrently and throughout the years of molecular genetics instruction. The authors contend that students are able to understand the concepts included in the meiotic and genetic models to some degree before understanding the molecular model in its entirety. Student progress, then, would be described as a deeper understanding of each of the three conceptual models and how they are related (reviewed in Duncan, 2011). Teaching the three models concurrently in increasing sophistication over time is consistent with the views of Stewart et al. (2005) who explained that literacy in molecular genetics consists of understanding each of the three models and being able to integrate the three models into coherent explanations of genetic phenomena.

The main conceptual difference between the two molecular genetics learning progressions is the placement of the three molecular genetics conceptual models in relation to the classroom instruction. Roseman et al. (2006) argued that the molecular model should be taught first in the early grades because understanding both the genetic
### Table 1

**Molecular Genetics Learning Progression Created by Duncan et al. (2009)**

<table>
<thead>
<tr>
<th>Components of Big Idea</th>
<th>Level 1: Grades 5–6</th>
<th>Level 2: Grades 7–8</th>
<th>Level 3: Grades 9–10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question:</strong> How do genes influence how we, and other organisms, look and function? Big Idea: All organisms have genetic information that is universal and specifies the molecules that carry out the functions of life. While all cells have the same information, cells can regulate which information is used (expressed).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(A)</strong> All organisms have genetic information that is hierarchically organized</td>
<td>Humans, animals, plants, fungi, and bacteria have genes (genetic information) in their cells</td>
<td>The genetic information is found in the chromosomes of cells. Most sexually reproducing organisms have two sets of chromosomes. All cells of an organism have the same two chromosomal sets (except sex cells)</td>
<td>Genes are nucleotide sequences within the DNA molecule. DNA molecules make up chromosomes that make up our genome</td>
</tr>
<tr>
<td><strong>(B)</strong> The genetic information contains universal instructions that specify protein structure</td>
<td>Genes are instructions for how organisms grow, develop, and function</td>
<td>Genes are instructions for molecules (many of which are proteins) that carry out functions within the organism. All organisms use the same genetic language for their instructions</td>
<td>The genetic code is translated into a sequence of amino acids that makes up the protein. Almost all organisms use the same genetic code</td>
</tr>
<tr>
<td><strong>(C)</strong> Proteins have a central role in the functioning of all living organisms and are the mechanism that connects genes and traits</td>
<td>Cells have to carry out many essential functions to live. Within cells organelles do specific functions. The structure of cells, tissues, and organs determines their function. Our body has multiple levels of organization and changes at one level may affect another</td>
<td>Proteins are like little machines that do the work of the cell. Proteins have shapes and properties that afford their functions. There are different types of proteins (enzymes, receptors, etc.) Changes to genes can result in changes to proteins, which can affect the structures and functions in the organism</td>
<td>Proteins have particular three-dimensional shape determined by their amino acid sequence. Proteins have many different kinds of functions that depend on their specific properties. There are different types of genetic mutations that can affect the structure and thus function of proteins and ultimately the traits</td>
</tr>
<tr>
<td><strong>(D)</strong> All cells have the same genetic information but different cells use (express) different genes</td>
<td>Different cells have some common and some different structures and functions</td>
<td>Different cells have different repertoires of proteins. Proteins carry out the basic (“housekeeping”) and unique functions of the cell</td>
<td>All cells have the same genetic content, but what genes are used by the cell (expressed) is regulated</td>
</tr>
<tr>
<td><strong>Question:</strong> Why do we, and other organisms, vary in how we look and function? Big Idea: There are patterns of gene transfer across generations. Cellular and molecular mechanisms drive these patterns and result in genetic variation. The environment interacts with our genetic makeup leading to variation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(E)</strong> Organisms reproduce by transferring their genetic information to the next generation</td>
<td>All organisms reproduce and transfer their genetic information to their offspring. Cells divide to make new cells each with all the genetic information. In larger organisms each parent contributes half the genetic information to the new generation</td>
<td>Before cells divide the chromosomes sets are duplicated and then two new cells are formed each with two chromosomal sets. In sexually reproducing organisms chromosome sets are randomly assorted into gametes through the process of meiosis (one full set in each sex cell). This process creates sex cells that have only one set of chromosomes</td>
<td>DNA replication is tightly regulated to prevent errors. During the process of meiosis chromosomes can swap sections and create new combinations of gene versions on a given chromosome, which creates more genetic variation</td>
</tr>
</tbody>
</table>
Understanding the hierarchical organization of the genetic material is also important as it provides connections between the molecular and meiotic models. Specifically, each DNA molecule is folded and packed into a larger structure called a chromosome—the basic structure that is passed down from one generation to the next (meiotic model). Genes are segments of the DNA molecule that each make-up a single information unit, or sentence in the genetic language (code for a single functioning molecule). Nucleotides are the building blocks of the DNA molecule and represent letters in the genetic language. We know that students have difficulties understanding the relationships between DNA, nucleotides, genes, and chromosomes (Lewis & Wood-Robinson, 2000) and that without understanding the relationships between these genetic structures students may not be able to coordinate the molecular and meiotic models. For example, explaining why some traits are usually inherited together entails understanding that the genes for these traits are located on the same chromosome.

### Components of Big Idea

<table>
<thead>
<tr>
<th>Level 1: Grades 5–6</th>
<th>Level 2: Grades 7–8</th>
<th>Level 3: Grades 9–10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F) There are patterns of correlation between genes and traits and there are certain probabilities with which these patterns occur</td>
<td>We vary in how we grow and function. For a given trait there are variations. Different organisms have different versions of the trait</td>
<td>Individuals have two versions for each gene (alleles). Each chromosome in the set carries one version of the gene. There are patterned correlations between the variants of the genes and the resulting trait</td>
</tr>
<tr>
<td>(G) Changes to the genetic information can cause changes in how we look and function (phenotype), and such variation in the DNA can serve as a way to identify individuals and species</td>
<td>Different organisms vary in how they look and function because they have different genetic information. Even within a group of organisms there is variation in traits</td>
<td>The genetic information can sometimes change. Changes in the genetic information can result in changes to the structure and function of proteins. Some changes can be beneficial, others harmful, and some neutral to the organism in its environment. Chromosomes, like X and Y, also vary in boys versus girls</td>
</tr>
<tr>
<td>(H) Environmental factors can interact with our genetic information</td>
<td>The environment can affect our traits. Even organisms that are related may end up looking or behaving differently</td>
<td>The environment can influence cell function through changes at the protein level (type and amount)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Environmental factors can cause mutations in genes, or alter gene expression</td>
</tr>
</tbody>
</table>

and meiotic models are dependent on understanding the molecular model. The authors also stated that the meiotic model should be taught next and the genetic model taught last in the later grades because the genetic model is the most abstract model of the three. Duncan et al. (2009) posited that all three models should be taught concurrently because understanding molecular genetics includes reasoning in and across all three models, consistent with Stewart et al.’s (2005) views on molecular genetics literacy. Duncan et al. (2009) stated that simplified models of each of the concepts should be introduced in the early grades and then built upon over the later grades.

The discrepancy between the two learning progressions is due to gaps in molecular genetics research. While several studies have examined students’ understandings of the meiotic and genetic models (e.g. Buckley et al., 2004; Cartier & Stewart, 2000; Jungck & Calley, 1985; Tsui & Treagust, 2003), there is a lack of research about how well middle school and high school students are able to reason about the molecular model. Few studies have been aimed at supporting and examining student understandings of the molecular model (Duncan & Tseng, 2011; Gelbart & Yarden, 2006; Rogat & Krajcik, 2006) and very few studies have been aimed at supporting students learning all three models concurrently (Cartier & Stewart, 2000; Duncan, Castro, & Bhojraj, 2013; Freidenreich et al., 2011). The lack of research in these areas makes ruling out either one of the proposed learning progressions very difficult (reviewed in Duncan, 2011). A recent paper by the Duncan lab (Freidenreich et al., 2011) does support the claim that in their learning progression students are able to reason, to some extent, in all three models concurrently. They found that middle school students’ (grades 6-8) understanding of the three models in molecular genetics grew with significant
learning gains and a large effect size in each model. The authors noted that even though student reasoning did not progress as much as they would have hoped in some areas, middle school students were able to reason to some extent in all three models concurrently and that a more targeted curriculum designed to promote reasoning among the three models should help students understand how the three models of molecular genetics are inter-related. The Duncan lab also recently documented that a middle school biology course which sequenced instruction of the molecular model prior to instruction of the genetic and meiotic models demonstrated increased student achievement in the models compared to students in a biology course which sequenced instruction of the molecular model after instruction of the genetic and meiotic models (Duncan et al., 2013). Clearly more research is needed before an empirically derived scope and sequence for molecular genetics instruction can be recommended.

Both the Roseman et al. (2006) and Duncan et al. (2009) molecular genetics learning progressions span grades 5-10 but do so in a different visual format. The progression developed by Roseman et al. (2006) takes the format of Project 2061’s Atlas of Science Literacy (AAAS, 2001). Although the progression was adapted from strand maps in the Atlas which contain grade bands, the progression itself does not clearly indicate what benchmarks should be reached by which grade (Figure 1). The progression by Duncan et al. (2009) is visually represented by a table (Table 1) which is organized around two questions: how do genes influence how we, and other organisms, look and function? and why do we, and other organisms, vary in how we look and function? The table includes questions broken down into eight “Big Ideas” with a learning performance for three different levels for each of the “Big Ideas.”
From a practical standpoint, the learning performances described in the progression by Duncan et al. (2009) are more useful for teachers and researchers because they are divided into levels and they indicate what learning performance could be expected of the students by certain grades. On the progression by Roseman et al. (2006), it is unclear how quickly students should progress through the ideas and where students could be at certain grades; however, teachers may be more familiar with the strand map layout since it mirrors those seen in the *Atlas of Science Literacy* (AAAS, 2001).

Since both progressions are both relatively new and theoretical, neither have been fully empirically validated. The Roseman et al. (2006) progression has yet to be empirically tested by any research group; the Duncan lab has begun to empirically test their progression (Freidenreich *et al.*, 2011, Shea & Duncan, 2013). The paper by Freidenreich *et al.* (2011) described results of testing the middle school portion of the progression in only one classroom context. The paper by Shea & Duncan (2013) built upon Freidenreich *et al.*’s (2011) research findings and proposed revisions and refinements to two of the eight constructs: B and C. Additionally, a paper by Shea, Duncan, & Giannetti (2013) discussed inclusion of a fourth level for each of the constructs but did not outline learning performances for the fourth levels. The authors also discussed the benefits of placing curricular emphasis on certain constructs that appear to be more relevant to mainstream scientific issues and decisions (such as constructs A, B, C, and G) despite the Duncan *et al.* (2009) progression placing equal emphasis on each of the eight constructs. Both progressions also lack information about classroom instruction and assessments. As these learning progressions are empirically tested and modified according to the data obtained, more instructional supports and
resources will be added to the progressions as well. The instructional materials and assessments will be resources that have been shown to support the progression of students through the levels of the learning progression (Duncan & Hmelo-Silver, 2009). Adding instructional and curricular supports contribute to more practical and useful progressions that support students’ understandings of molecular genetics.

**Purpose for Study**

In the past several decades there has been a large push for increasing scientific literacy (AAAS, 1989; AAAS, 1993; NRC, 1996; NRC 2012), especially in the areas that are rapidly advancing, like molecular genetics. Literacy in molecular genetics is especially important because the general public is beginning to encounter molecular genetics during the course of their everyday lives and this content area will likely have the most immediate and direct impact on an individual of any science area. Much research has been done on student understandings of molecular genetics and the consensus is that the concepts are difficult both to learn and teach (Fisher, 1992; Friedrichsen & Stone, 2004; Horwitz, 1996; Kindfield, 1992; Lewis & Kattmann, 2004; Lewis & Wood-Robinson, 2000; Marbach-Ad & Stavy, 2000; Stewart et al., 2005; Stewart & Van Kirk, 1990; Venville & Treagust, 1998; Wynne et al., 2001). Additionally, research repeatedly indicates that the general public is making uninformed decisions based on their lack of literacy in molecular genetics (Fisher, 1992; Garton, 1992; Kindfield, 1992) and that many high school graduates are even ill-equipped to make informed decisions about topics in molecular genetics despite passing the required life science courses (Lanie et al., 2004; Lewis & Wood-Robinson, 2000).
Two learning progressions in the molecular genetics content area have been produced (Duncan et al., 2009; Roseman et al., 2006); however, both progressions are hypothetical as neither have been fully empirically tested. The Roseman et al. (2006) progression has yet to be empirically tested by any research group; the Duncan lab has begun to empirically test their progression (Freidenreich et al., 2011, Shea & Duncan, 2013). To date, the Duncan lab has only published results of testing the middle school portion of their progression in only one classroom context and published revisions and refinements of two of the eight constructs in their progression. Empirical studies of all learning progressions lead to revisions and refinement based on data obtained. These studies also eventually document instructional and curricular support in the form of instructional materials and assessments that have been shown to support the progression of students through the levels of the learning progression (Duncan & Hmelo-Silver, 2009). Validation of learning progressions through empirical studies and the addition of instructional and curricular support make these learning progressions more practical and useful for researchers, as well as teachers, to help support students’ understandings.

This research study will fill several gaps in molecular genetics research, based on the following research questions:

• **RQ1**: Where do high school students in different classroom contexts appear on the Duncan *et al.* (2009) molecular genetics learning progression?

• **RQ2**: What impact do intervention units have on learning performances in different classroom contexts?

• **RQ3**: How can the molecular genetics learning progressions be revised and refined based on empirical testing in different classroom contexts?
First and foremost, this study is designed to empirically test the Duncan et al. (2009) molecular genetics learning progression using 10th grade students in three different classrooms in two different schools to see where students appear on the progressions. The learning progressions remain hypothetical models of student learning until they are empirically tested and validated through multiple iterative rounds of empirical testing. The empirical tests of learning progressions must be conducted in several different contexts. This research will test the upper bounds of the progression by using 10th graders in three different classroom contexts.

It is important to note, however, that due to the previously described general lack of molecular genetics literacy, it is expected that 10th grade students will appear on the extremely low levels of achievement and learning performances before classroom instruction during the 10th grade year. Although the study focuses on one year of classroom instruction, it is expected that students will progress from the low learning performances at the beginning of the year to the higher learning performances as the molecular genetics instructional period continues. Tracking student progress through the entire learning progression will be documented.

This study will help determine how well high school students are able to reason about the molecular model, another gap in current molecular genetics research. It will also help determine the extent to which students are able to concurrently reason between all three models, given that 10th grade students should be able to reason at the upper limits of the progressions. Currently, only two studies (Cartier & Stewart, 2000; Freidenreich et al., 2011) have engaged students in learning all three models concurrently, however in the Cartier & Stewart (2000) study, instruction on the molecular
model was weak and the researchers failed to assess the students’ abilities to develop explanations using the molecular model. The focus of the Freidenreich et al. (2011) study was determining how well middle school students were able to reason in and between the three models, but some limitations to the study was that it only focused on one classroom context and did not probe student ideas in each of the eight “Big Idea” constructs of the Duncan et al. (2009) progression. Although this specific study is more concerned with the molecular model, assessment questions have been designed to probe each of the eight constructs of the Duncan et al. (2009) learning progression, and thus, all three of the models in three different classroom contexts.

Empirically testing the progression in classrooms will provide valuable information that will help revise and refine both the Duncan et al. (2009) and Roseman et al. (2006) molecular genetics learning progressions. While this research does not aim to validate one progression over the other, it will provide data to revise and refine the ideas in the progressions and allow the progressions to be merged into a single progression which encompasses ideas included on both progressions.

The research will also determine how intervention units that target specific aspects of the Duncan et al. (2009) progression impact student learning performances in molecular genetics. Differing from traditional instruction, the intervention units introduce proteins and their functions before addressing DNA and its structure, as both then Roseman et al. (2006) and Duncan et al. (2009) learning progressions have suggested. The intervention units strive to help students better understand molecular genetics by specifically addressing “Big Idea” constructs B, C, D, F, G, and H from the Duncan et al. (2009) learning progression (Table 1) in an inquiry-based environment.
Well-developed learning progressions have been both empirically validated and contain instructional materials and assessments that have been shown to support the progression of students through the levels of the learning progression (Duncan & Hmelo-Silver, 2009). Neither molecular genetics learning progression has yet been fully validated and neither contains any instructional or curricular supports that have been shown to promote student progression through the learning progression. This study will determine the impact of the three intervention units developed by the author in relation to student progress through the learning progressions. If the units and assessments are found to help support student progress, they could then be added to the progressions. As instructional and curricular supports are added to the progressions, the learning progressions become more practical and useful for teachers and researchers.

It is hypothesized that students will hold ideas not included in the molecular genetics learning progressions, the empirical data obtained in this study can be used to revise and refine the progressions, and students who complete the activities in the intervention units will achieve higher learning performances in constructs targeted by the units than the students who do not complete the intervention units.
II. Methods

Study Context

The study was conducted during the 2011-2012 school year in three 10th grade biology classrooms in two different schools (Table 2). School 1 is a suburban public school (grades 6-12) with a STEM focus. The school’s academic performance rating for that school year was designated as Excellent. The student population is approximately 36% black, hispanic, asian/pacific islander, or multi-racial and 64% white. Approximately 25% of all students were considered economically disadvantaged and 8.6% had disabilities. The building’s poverty status is considered low poverty. School 2 is an urban public school (grades 7-12) with an arts focus. The school’s academic performance rating for the previous school year was designated as Effective. The student population is approximately 66% black, hispanic, or multi-racial and 33% white. Approximately 56% of all students were considered economically disadvantaged and 11.4% has disabilities. The building’s poverty status is considered medium-high poverty.

School 1 had one 10th grade biology teacher and School 2 had two 10th grade biology teachers. The three teachers, each with their own classroom and their own students, were followed in this study (Table 2, contexts 1-3). The teacher in context 1 was in her second year of teaching, both of which had been at School 1. She has a Master of Science in Education (Secondary Science) and participated in a small pilot study teaching the intervention units during the previous school year. The teacher reported that her instructional strategies a high percentage of activity-based and inquiry
Table 2

*Classroom Contexts Used For Study*

<table>
<thead>
<tr>
<th>Category</th>
<th>Context 1</th>
<th>Context 2</th>
<th>Context 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Environment</td>
<td>Suburban public, School 1</td>
<td>Urban public, School 2</td>
<td>Urban public, School 2</td>
</tr>
<tr>
<td>Grades in School</td>
<td>6-12</td>
<td>7-12</td>
<td>7-12</td>
</tr>
<tr>
<td>School Focus</td>
<td>STEM</td>
<td>Arts</td>
<td>Arts</td>
</tr>
<tr>
<td>Demographics</td>
<td>16.4% black non-hispanic, 4.2% hispanic, 11.1% multi-racial, 3.8% asian/pacific islander, 64.2% white non-hispanic</td>
<td>58.8% black non-hispanic, 2.6% hispanic, 4.7% multi-racial, 33.2% white non-hispanic</td>
<td>58.8% black non-hispanic, 2.6% hispanic, 4.7% multi-racial, 33.2% white non-hispanic</td>
</tr>
<tr>
<td>Economically Disadvantaged</td>
<td>24.8%</td>
<td>55.5%</td>
<td>55.5%</td>
</tr>
<tr>
<td>Students with Disabilities</td>
<td>8.7%</td>
<td>11.4%</td>
<td>11.4%</td>
</tr>
<tr>
<td>School Academic Performance Rating</td>
<td>Designation: Excellent; 16/16 State Indicators Met; Performance Index: 105.4/120; AYP met</td>
<td>Designation: Effective; 14/17 State Indicators Met; Performance Index: 98.1/120; AYP not met</td>
<td>Designation: Effective; 14/17 State Indicators Met; Performance Index: 98.1/120; AYP not met</td>
</tr>
<tr>
<td>Teacher</td>
<td>Ms. Clark</td>
<td>Mrs. Robinson</td>
<td>Ms. Smith</td>
</tr>
<tr>
<td>Highest Degree Completed</td>
<td>MS Ed. (Secondary Science)</td>
<td>Master of Science Teaching (MST)</td>
<td>Bachelor’s</td>
</tr>
<tr>
<td>Years Teaching</td>
<td>2</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>Years Teaching Biology</td>
<td>2</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Years Teaching Biology at Current School</td>
<td>2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Approx. Percentage of Activity-Based Learning</td>
<td>75-80%</td>
<td>60%</td>
<td>95%</td>
</tr>
<tr>
<td>Approx. Percentage of Inquiry Activities</td>
<td>80%</td>
<td>60%</td>
<td>5%</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Category</th>
<th>Context 1</th>
<th>Context 2</th>
<th>Context 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Genetics Sequence Taught</td>
<td>cells, cell specialization (Unit 1), infectious disease, biochemistry/proteins (Unit 2), energy &amp; transport, DNA, cell division/cancer, genetic disease (Unit 3), genetics</td>
<td>protein synthesis, chromosome/DNA/cell reproduction, cell specialization (Unit 1), protein structure &amp; function (Unit 2)</td>
<td>structure &amp; function of cell, homeostasis &amp; transport, photosynthesis/respiration, cell reproduction, fundamentals of genetics, nucleic acids &amp; protein synthesis, gene expression, inheritance patterns &amp; human genetics</td>
</tr>
<tr>
<td>Intervention Units Completed</td>
<td>3</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Class Periods of 10th Grade Biology</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Total 10th Grade Biology Students</td>
<td>62</td>
<td>65</td>
<td>86</td>
</tr>
<tr>
<td>Students Agreeing to Participate in Study</td>
<td>55</td>
<td>27</td>
<td>39</td>
</tr>
<tr>
<td>Students Interviewed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Interview</td>
<td>30</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Middle Interview</td>
<td>30 (26)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Post-Interview</td>
<td>26 (23)</td>
<td>22 (22)</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note.* Teacher Names are pseudonyms; Years Teaching include year of research study; Percentage of Activity-Based Learning, Percentage Inquiry Activities, and Molecular Genetics Sequence Taught are teacher-reported; Total 10th Grade Biology Students is number of students enrolled at beginning of the school year; Students Agreeing to Participate in Study is number of students consenting to written work collection; Students Interviewed is number of students interviewed for each interview with number in parentheses representing number of original students interviewed for each interview.
learning (~80%) and the teacher completed all three intervention units during the course of the molecular genetics instructional period. The teacher in context 2 was in her 25th year of teaching, 12 of which had been at School 2. She has a Master of Science Teaching degree and was the researcher’s cooperating teacher during the previous school year. The teacher reported that her instructional strategies included a moderate percentage of activity-based and inquiry learning (60%) and noted that she considered activity-based and inquiry learning to be the same. Due to time constraints with state testing, this teacher completed the first intervention unit in its entirety and did a shortened version of the second unit in three days which was suggested by the researcher to include only activities 2, 3, and 5. The teacher in context 3 was in her second year of teaching and her first year of teaching at School 2. She has a Bachelor’s degree and reported that her instructional strategies included a very large percentage of activity based learning (95%) and a very small amount of inquiry activities (5%). It is worth noting that these percentages add up to 100%, possibly indicating that this teacher may not fully understand the science education definitions of activity and inquiry learning. This teacher did not receive the intervention units and taught molecular genetics with district-supplied curriculum and her own materials. The teacher-reported sequence of molecular genetics topics taught in their classrooms is also included in Table 2.

This study is not a comparison study between classrooms and/or teachers; the goal is to map student learning in three different contexts to the molecular genetics learning progressions. Students from each of the three classroom contexts were included in the study (Table 2). Written assessments were collected from all students who consented to written work collection (n = 121) in all three classrooms. Interviews were conducted.
with students from contexts 1 \((n = 34)\) and 2 \((n = 24)\). One student moved between classrooms in School 2, going from context 2 to context 3, towards the beginning of the year; this student was included in the student count for context 3 because the student only completed half of the first intervention unit and spent the majority of her molecular genetics instructional period in context 3. A more detailed break-down of student interviews is included in Table 2.

Student understandings of molecular genetics were studied before, during, and after instruction of the relevant concepts and intervention units. The teachers taught molecular genetics in their own teaching styles, using the resources and lessons they chose. Two of the classrooms (one at each school; contexts 1 & 2) received the intervention units, while the third (context 3) did not. The teachers that received the intervention units used the units as a supplement to their instruction or replaced some lessons they had typically done in the past. In each context, the intervention units were implemented during the course of the regular molecular genetics instructional period, which varied among teachers. The molecular genetics instructional period for 10th grade typically consists of cells, homeostasis and transport, photosynthesis and cellular respiration, cellular differentiation, DNA and its structure, the genetic code, transcription and translation (the central dogma), structure-function relationships in proteins, genetics, heredity, and mutations and genetic disease. The teacher-reported sequence of molecular genetics topics taught in their classrooms is included in Table 2.

**Intervention Units**

Three intervention units were created by the researcher to provide targeted instruction and curriculum to several of the constructs in the Duncan *et al.* (2009)
progression. The units differ from traditional classroom instruction in two main ways. First, proteins and their functions inside of cells are discussed before DNA is introduced. Second, learning is targeted to specific constructs of the Duncan et al. (2009) molecular genetics learning progression. Although the progression has eight “Big Idea” constructs, the intervention units only focused on six: B, C, D, F, G, H (Table 1). These six constructs are mainly concerned with the molecular model of genetics and describe how proteins are important in cells and how changes to the DNA can affect protein and cell function. Each intervention unit takes approximately 6-7 instructional days to complete in its entirety.

The intervention units are different from traditional pedagogy because they were written as inquiry units centered around driving questions that provide purpose for the content learning. Each unit had its own driving question. The intervention units included scientific practices such as modeling, constructing evidence-based explanations, and argumentation from evidence that were included in the Framework and Next Generation Science Standards (NGSS); additionally the content included in the intervention units aligned with the NGSS (HS-LS1-1, HS-LS1-4, HS-LS3-1, HS-LS3-2, HS-LS4-2).

The first intervention unit describes cellular differentiation; centers around the driving question “How do cells become cancerous?”; and includes ideas from constructs C, D, and H on the Duncan et al. (2009) learning progression. Students learn that all cells (except sex cells) contain the same DNA but express different parts to make the proteins necessary for their specific functions by examining RNA expression data (simplified to number of proteins) in differentiating cells. They learn about the functions of some different proteins and interpret data to learn that, for example, crystallin (a clear
protein that allows light to pass through unobstructed) is only expressed in cells in the eye. Students also explore stem cells and learn that they have not yet differentiated. Next they compare stem cells to cancer cells and learn that cancer cells are de-differentiated cells. The two classrooms that received the intervention units (contexts 1 and 2) completed this unit in its entirety.

The second intervention unit describes protein structure/function and how genes give observable traits, centers around the driving question “Why is a Siamese cat colored the way it is?”, and includes ideas from constructs C and G on the Duncan et al. (2009) learning progression. Students learn about structure and function of a variety of different proteins and learn that they are strongly correlated. They learn that heat and acid can denature proteins so that they can no longer function by doing a lab with Jell-O and fresh and concentrated fruit juices. Students then learn about the browning of fruit by observing apples in lemon juice or potatoes with catechol and relate this reaction to the reaction in a Siamese cat to give it the dark fur color. The main goal of this unit is for students to understand how proteins are the link between genes and traits. Context 1 completed this intervention unit in its entirety. Context 2 completed a condensed version of this unit in three days due to time constraints and only completed activities 2, 3, and 5 as suggested by the researcher; the selected activities highlight how proteins link genes and traits.

The third intervention unit describes DNA and protein mutations; centers around the driving question “Can we engineer a superhuman?”; and includes ideas from constructs B, C, F, and G on the Duncan et al. (2009) learning progression. Students learn about different DNA mutations and their effect on proteins with examples of
different genetic diseases. They learn how cells repair DNA damage and mutations and learn that different mutations can change proteins in different ways. Students then learn that Mendelian relationships can be explained at the molecular level as interactions between proteins. They also learn about increasing athletic performance with recombinant erythropoietin and are asked to genetically engineer a superhuman, specifically explaining how this would be done at the gene, protein, cell and trait level. The “gene, protein, cell, trait” scaffold was created by Duncan et al. (2011) for intervention units targeting middle school students. Only context 1 completed this unit and did so in its entirety.

Data Collected

Four sources of data were collected: written assessments, interviews, classroom video, and student written artifacts/worksheets. Written assessments developed by the researcher consisted of a pre- and post-test (Appendix A) and was administered to the students before beginning the molecular genetics units (pre-test) and after the molecular genetics units (post-test). The pre- and post-test consisted of the same test questions and only had one minor formatting change between the two in order to directly compare changes in student thinking.

The written assessment consisted of 15 questions, some with short scenarios, that required students to select a multiple choice answer and then explain in an open response question why they chose their answer. Each multiple choice question contained three responses from which students could choose. Each question was targeted to a specific construct in the Duncan et al. (2009) molecular genetics learning progression and the three multiple choice responses aligned with the three levels in the learning progression.
This format is similar to ordered multiple choice questions and item response theory (Adams & Wilson, 1992, 1996, Briggs, Alonzo, Schwab & Wilson, 2006). Using this format makes use of the distinct learning performances known for each construct. The 6 constructs discussed in the intervention units have at least two questions each (construct G has three questions) while the other 3 constructs have one question each in the assessment. The students in all three classrooms participated in the written assessments ($n = 121$).

Clinical interviews (Ginsberg, 1997) were conducted with students before, during, and after the intervention units. In context 1, 30 students were selected based on the teacher’s recommendations and represented all class periods, both genders, several ethnic backgrounds, and a range of abilities. All 24 students in context 2 who consented to interviews were selected for interviews. No students in context 3 were chosen for interviews since they did not implement the intervention units in class. The pre-interviews took place after the written pre-test assessment but before the start of the molecular genetics units. The middle interviews took place after the completion of the first two intervention units. The post-interviews took place after the written-post test assessment upon completion of the entire molecular genetics unit. Since Context 2 was unable to complete the third intervention unit, no middle interview was done after the second intervention unit; students in this classroom only participated in pre- and post-interviews. The interviews lasted a maximum of 35 minutes and students were asked to explain their answers on the pre- and post-test in more detail than they included on the assessment. Students were also asked to relate their answers and understanding of
molecular genetics to activities and lessons they had done in class. The interview protocol is provided in Appendix B.

In context 1, due to some students leaving the school (7 students) or not wanting to be interviewed again (1 student), 26 of the original 30 participated in the middle interview, and 23 participated in the post-interview. Four additional students who were not interviewed in the pre-interview participated in the middle interview to bring the total number of students interviewed in the middle interview to 30. The remaining 26 students who had been interviewed at some point were interviewed in the post-interview, 23 of which had completed all three interviews (Table 2). In context 2, one interviewed student left the school and another transferred out of the class leaving 22 participants in the post-interview. Since context 2 students only completed the pre- and post-interviews and all students who consented to interviews were initially selected, no additional students were interviewed in the post-interview (Table 2).

Classroom video was collected during seven target lessons in the intervention units. Video was collected in the two classrooms that received the intervention units and only one class period for each was videotaped. The seven target lessons were Unit 1: Activities 2, 4, 5; Unit 2: Activities 3, 5; Unit 3: Activities 2, 3. The lessons were targeted because they were key lessons to understanding the molecular genetics content and focused on specific aspects of the Duncan et al. (2009) learning progression. Although every attempt was made to video each of the target lessons in both classrooms, context 1 was videotaped for 5 lessons and context 2 was videotaped for 3 lessons. The discrepancy was caused by timing of the class periods to be videotaped, a teacher starting the units without researcher notification, and context 2 only completing 1.5 of the units.
The video documented the concepts taught, the ways in which the lessons were presented to the students, and the whole class discussions. The video also documented small group work and discussions the students had with one another while working through the intervention units. The classroom video provides evidence of student learning during the lessons and was useful to triangulate findings from the interviews and written assessments. The interviews prompted students to discuss events from class that aided them in their knowledge of the content. The classroom video provided evidence of the events that aided student learning in specific situations.

Student written artifacts were collected from the students in contexts 1 and 2. The written artifacts were the intervention worksheets which asked students to draw models, make predictions, and write explanations of phenomena in molecular genetics. The written artifacts were also useful to triangulate findings from the interviews and written assessments. They also provided a written record of how students were thinking through the concepts during the lessons and how their thinking changed over the course of the lessons.

**Theoretical Framework**

The theoretical framework used to analyze data in this project is based on several different theories in molecular genetics. First and foremost, student achievement in molecular genetics was mapped to the theoretical learning progression proposed by Duncan *et al.* (2009) and then correlated with the Roseman *et al.* (2006) learning progression. The Stewart *et al.* (2005) theory that molecular genetics literacy consists of understanding and integrating three inter-related conceptual models is also an important theory used in data analysis. Additionally, Duncan & Reiser’s (2007) “hybrid
hierarchical” structure of molecular genetics was an important aspect of the intervention units. Building upon the idea that an information level (genes) codes for a biophysical layer (proteins, cells, tissues, organs) and upon the suggestion by both molecular genetics learning progressions (Duncan et al., 2009; Roseman et al., 2006), the intervention units were designed to first introduce the biophysical layer and how proteins bring about observable traits before introducing the information level and the concept that genes simply code for proteins.

**Development of Coding Schemes**

Coding schemes based on the Duncan et al. (2009) learning progression were developed for each of the “Big Idea” constructs in the progression based on preliminary data obtained the previous school year and revised according to the larger current data set. Tables 3.1-3.9 show the construct “Big Idea” and question(s) developed to probe understanding of each big idea followed by the coding scheme and examples of student responses for each code. In each coding scheme, Levels 1, 2, and 3 correspond to the three levels in the Duncan et al. (2009) learning progression. The levels were given codes of 1, 2, and 3, respectively. The response to each of these levels corresponds to the learning performance outlined for each level in the progression. Because several intermediate ideas had been found from a preliminary study and in the larger data set, each coding scheme also contains levels between levels 1, 2, and 3. In all the coding schemes level 0 corresponds with no response or no idea or a response of “guess.” The other intermediate ideas were coded according to the already identified levels and given intermediate levels (such as 1+, 3-) and codes (such as 1.33, 2.67) based on where responses appear relative to the Duncan et al. (2009) described learning performances.
Construct A describes the idea that genetic information is hierarchically organized from nucleotides making up genes in the DNA and DNA making up chromosomes that make up our genome inside of each cell. Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.1.

Construct B describes the idea that genes simply code for the order of amino acids in a protein. Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.2.

Construct C describes the ideas that proteins have a central role in cellular functions and that they are the connection between genes and traits. Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.3.

Construct D describes the idea that all cells (except gametes) in an organism contain the same genetic information, but express different genes to produce the proteins necessary for the specialized cell’s specific function. Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.4.

Construct E describes the ideas that organisms transfer genetic information to their offspring, sexually reproducing organisms produce gametes containing half the genetic material by meiosis, and during meiosis chromosomes can swap sections (recombination) to produce further genetic variation in offspring. Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.5.
Table 3.1

*Codes Developed for Construct A*

**Big Idea**

All organisms have genetic information that is hierarchically organized.

**Question**

1. Put the following terms in some sort of order or pattern: DNA, gene, chromosome, nucleotide/base, cell, genome

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No correlation between words</td>
<td>1027 (pre): “Cell, chromosome, DNA, gene, genome, nucleotide/base. I found it easy to organize in alphabetical order.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>All incorrect correlations between words</td>
<td>1040 (pre): “Genes consist of genomes based upon handed down DNA which consists of cells made out of chromosomes made with nucleotide which is like the base of the whole thing.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Correct correlation between 2 words</td>
<td>1061 (pre): “I didn’t really know what order they went in but I knew gene was before DNA, and cells &amp; chromosomes were small so I put them near the beginning. I put DNA at the end because it seemed the only one that contained others in it.”</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Correct correlation between 3-4 words</td>
<td>1053 (post): “Because nucleotide/base make up chromosomes, which then make up DNA. Genes are constructed of DNA and are in our cells.”</td>
</tr>
<tr>
<td>2+</td>
<td>2.33</td>
<td>Correct correlation between 5 words</td>
<td>1017 (post): “Because nucleotides and bases make up DNA which codes for a gene which makes up a genome. The genome goes inside the chromosome which goes in a cell.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Correct correlation between all 6 words</td>
<td>1018 (post): “The nucleotide/bases make up genes, genes are a part of the DNA, DNA is compacted into chromosomes, all chromosomes and DNA make up a genome, and every cell has the genome for the entire organism.”</td>
</tr>
</tbody>
</table>
Table 3.2

*Codes Developed for Construct B*

### Big Idea

The genetic information contains universal instructions that specify protein structure.

### Questions

1. Check the box next to the statement you think best explains how DNA is involved in muscle function.

2. Check the box next to the statement that you think best explains why DNA is sometimes called the “genetic code.”

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>no mention of function</td>
<td>1011 (pre): “I guessed.”</td>
</tr>
<tr>
<td>0+</td>
<td>0.33</td>
<td>DNA is passed down or is your genes or tells about your genes (non-informational in nature)</td>
<td>1021 (pre): “Because genes are passed down from parents to children.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>DNA is your identification code or is instructions or codes for genes (informational in nature)</td>
<td>1003 (pre): “DNA is the base for all instructions.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>DNA contains instructions to tell your body how to grow, function, and develop</td>
<td>1009 (pre): “I know that DNA is the blueprints for our entire body.”</td>
</tr>
<tr>
<td>2-</td>
<td>1.67</td>
<td>DNA has instructions for cells or organs or tissues</td>
<td>1043 (post): “Because this answer explains how DNA is able to function properly because of the genes. The muscle cells are only able to function properly with the correct gene.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>DNA codes for molecules or amino acids or proteins that carry out functions</td>
<td>1035 (post): “Because DNA does not code for instructions for the body. Just to make proteins.”</td>
</tr>
<tr>
<td>2+</td>
<td>2.33</td>
<td>DNA codes for proteins, proteins are made of amino acids</td>
<td>1051 (post): “These genes do code for proteins and the amino acid sequences are what make up proteins.”</td>
</tr>
<tr>
<td>3-</td>
<td>2.67</td>
<td>DNA codes for amino acids which make up proteins</td>
<td>1021 (post): “The DNA codes for the specific amino acids which make up the protein. The protein performs the functions.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>DNA is translated into a sequence of amino acids that make up a protein</td>
<td>1054 (post): “It explains that genes code for proteins by giving the order of amino acids that make up the protein and gave the effects of a protein on a cell.”</td>
</tr>
</tbody>
</table>
### Table 3.3

_Codes Developed for Construct C_

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Genes and traits are not connected.</td>
<td>1039 (pre): “Because it sounded good.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>Changes in genes can lead to changes in traits, genes give traits, changes instructions</td>
<td>1024 (pre): “Genes carry instructions so it would be a change to the instructions not like the protein supply.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Changes in genes can lead to changes in cells, genes tell cells what to do to give traits</td>
<td>1023 (post): “The gene starts it all, any change in the gene will change the cell.”</td>
</tr>
<tr>
<td>2-</td>
<td>1.67</td>
<td>Changes to genes can change proteins, genes make proteins</td>
<td>1041 (post): “The gene does change dystrophin protein.”</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Changes to genes can result in changes to proteins which can affect structure and function in an organism, genes make proteins which give traits (no mention of how proteins give traits)</td>
<td>1011 (post): “Because it talks about the process of how the gene changes the protein which changes the function because the proteins give instructions for cell’s function.”</td>
</tr>
<tr>
<td>2+</td>
<td>2.33</td>
<td>Changes to genes change amino acids</td>
<td>1017 (post): “The amino acids change which will change the function of the muscle fiber.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-</td>
<td>2.67</td>
<td>Changes in genes change the amino acid to change the protein to give the trait</td>
<td>1024 (post): “A change to the gene will change the amino acids which will change the protein and then change the cell and then make the trait we see.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Different types of genetic mutations can affect structure and function of proteins and ultimately traits, genes give proteins which have a function to give trait</td>
<td>1048 (post): “This student explained how flaws in the gene change the amino acid sequence which changes the function of the protein.”</td>
</tr>
</tbody>
</table>
Table 3.4

*Codes Developed for Construct D*

**Big Idea**

All cells have the same genetic information but different cells use (express) different genes.

**Questions**

1. The dystrophin gene is involved in anchoring muscle fibers. For each cell type (muscle and blood), place an “X” in each box where you think the dystrophin gene, mRNA, or protein are present.

2. Which student do you think best explained why skin and nerve cells looked different?

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No idea why cells are different</td>
<td>1012 (pre): “I think that is the right answer.”</td>
</tr>
<tr>
<td>0+</td>
<td>0.33</td>
<td>Cells are different because they are in different places</td>
<td>2040 (post): “If it’s in the brain, it’s going to look probably totally different. If it’s on the skin, it’s probably going to be totally different... because it’s in different positions in your body.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>Cells are in different places so different things (i.e. air) can affect them</td>
<td>hypothetical: “Skin cells are exposed to the air and nerve cells are deep in your brain, so they have to look different.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Specialized cells are different because they have certain functions</td>
<td>1004 (pre): “Because usually when something looks different, it’s because it does something different.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1+</td>
<td>1.33</td>
<td>DNA tells your body what to do or makes you who you are or tells cells to look different or use different parts of DNA for different functions</td>
<td>1056 (pre): “Because I think the DNA makes up the cells and tells them how to form.”</td>
</tr>
<tr>
<td>2-</td>
<td>1.67</td>
<td>Specialized cells have different things inside of them to do function</td>
<td>1052 (pre): “The different cells have different proteins which to me it makes sense why it would look different because it has different stuff inside of it.”</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Specialized cells have different proteins inside of them to do function</td>
<td>1043 (post): “Proteins are special to their functions which can cause the cells to perform and look differently.”</td>
</tr>
<tr>
<td>2+</td>
<td>2.33</td>
<td>Cells have different DNA to give different proteins</td>
<td>1038 (mid): “[The cells] have different parts of the DNA to make up different proteins.”</td>
</tr>
<tr>
<td>3-</td>
<td>2.67</td>
<td>All cells have the same DNA but different proteins (no explanation of how this works)</td>
<td>1051 (mid): Different cells do have different proteins in them. Like neurons don’t have the same proteins as muscle cells... different cells do have the same DNA in them.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Specialized cells contain the same DNA but expression of genes/proteins is specific to that type of cell</td>
<td>1009 (post): “All cells have the same DNA but use different parts of the DNA to produce different proteins.”</td>
</tr>
</tbody>
</table>
Table 3.5

*Codes Developed for Construct E*

**Big Idea**

Organisms reproduce by transferring their genetic information to the next generation.

**Questions**

1. Check the box next to the statement you think best explains what a baby bunny would look like.

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No explanation</td>
<td>1002 (pre): “I used an educated guess.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>DNA/traits are passed down, organisms can only get traits of parents</td>
<td>1022 (pre): “Because the baby bunnies would share the traits of its parents.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>DNA is passed down, each parent contributes half to the next generation (evidence of parent giving half of DNA to offspring)</td>
<td>1001 (post): “Because 3 out of 4 fur colors are grey, making it a 75% chance. Also 2 out of 4 eye colors are black and the other 2 are blue. This creates a 50/50 chance with eye color.”</td>
</tr>
<tr>
<td>1+</td>
<td>1.33</td>
<td>Traits/alleles are randomly assorted independent of chromosomes, each sex cell contains only one allele</td>
<td>1024 (post): “There is a 50% chance that the bunny would have brown fur and a 50% chance the bunny would have grey fur. Brown is dominant over the grey. Same thing with the 75 and 25 chance that there would be black versus blue. Because black is dominant over blue... The fact that you have three recessive genes for the fur color gives you a chance at having 50-50. Because there’s only two recessive genes for the eye color, you’ve only got a 25% chance.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-</td>
<td>1.67</td>
<td>Chromosomes are randomly assorted and each sex cell contains only one set of chromosomes</td>
<td>1010 (pre): “There’s two things of genetics that they can inherit. So it can kind of mismatch and they could get any particular thing from either thing of genetics.”</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Cells go through meiosis to create gametes, chromosomes are randomly assorted and each sex cell contains only one set of chromosomes</td>
<td>hypothetical: “The female rabbit has a grey/black chromosome and a brown/blue chromosome for fur/eye color. The male rabbit has a grey/black chromosome and grey/blue chromosome for fur/eye color. This means there is a 50% chance the bunny will have grey fur and black eyes, a 25% chance the bunny will have brown fur and black eyes, and a 25% chance the bunny will have brown fur and blue eyes.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Chromosomes can swap sections and recombine during meiosis which creates more genetic variation</td>
<td>hypothetical: “The brown and blue alleles are on the same chromosome. Since brown is dominant over grey for fur color and blue is recessive to black for eye color, it would be extremely rare for a bunny to have grey fur and blue eyes. The only way it could happen would be if there was recombination during meiosis and the grey allele traded places with the brown allele so that a new chromosome had both recessive alleles together: grey and blue.”</td>
</tr>
</tbody>
</table>
Construct F describes the ideas that there are patterned correlations between the variants of genes and the resulting traits, individuals have two alleles for each gene which vary in terms of nucleotide sequence, and dominant and recessive relationships can be explained by the function and interaction of gene products (proteins). Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.6.

Construct G describes the ideas that changes to the genetic information can cause changes in phenotype, that these changes are necessary for evolution and natural selection, and variation in DNA can serve as a way to identify individuals and species. Construct G was divided into two different coding rubrics, the first describing genetic variation between individuals and species (G1, Table 3.7), and the second describing genetic changes and how they relate to evolution and natural selection (G2, Table 3.8). Questions, levels, codes, descriptions, and student response examples for these “Big Ideas” are described in Tables 3.7 and 3.8.

Construct H describes the idea that environmental factors can influence our genetic material by causing mutations to or altering expression of genes. Questions, levels, codes, descriptions, and student response examples for this “Big Idea” are described in Table 3.9.

Data Analysis

Student learning was mapped to the Duncan et al. (2009) molecular genetics learning progression using the previously described coding rubrics for each construct. For each of the “Big Idea” constructs, students were mapped to an initial level (using the pre-test and pre-interviews). Student progression over time through the molecular
Table 3.6

*Codes Developed for Construct F*

### Big Idea

There are patterns of correlation between genes and traits and there are certain probabilities with which these patterns occur.

### Questions

1. Check the box next to the statement you think best explains how Bill was able to produce pink-flowered snapdragons from crossing a white-flowered plant with a red-flowered plant.

2. Check the box next to the statement you think best explains what the colored bars in the DNA are.

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No explanation</td>
<td>1003 (post): “Best description, most ‘scientific’.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>There are different versions of traits and organisms can have these different versions</td>
<td>1038 (pre): “Because I think the plants changed traits.”</td>
</tr>
<tr>
<td>1+</td>
<td>1.33</td>
<td>Male traits are dominant or females look more like their mother or two traits can be mixed or one trait can “win” out</td>
<td>1031 (post): “When you have two different colors, you cross them to make another color.”</td>
</tr>
<tr>
<td>2-</td>
<td>1.67</td>
<td>Each chromosome carries one allele of a gene and the alleles with the greatest number are dominant or organisms have two genes for the same trait</td>
<td>1032 (post): “The red/white allele are different versions of the same allele, but are not the same gene.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>Each chromosome carries one allele of a gene, there are (correct) patterns which determine the resulting trait based on the alleles</td>
<td>1048 (post): “The incomplete dominance in the flower makes it pink. That comes from a red allele &amp; a white allele.”</td>
</tr>
<tr>
<td>2+</td>
<td>2.33</td>
<td>Alleles differ in nucleotide sequence or protein interactions give trait variations</td>
<td>1030 (post): “Plants do have changes in the DNA, they don’t always have the same between plants, this allows changes in color.”</td>
</tr>
<tr>
<td>3-</td>
<td>2.67</td>
<td>Alleles differ in nucleotide sequence affecting the protein which gives trait variation</td>
<td>1053 (post): “The gene has small changes in DNA that makes the pigment protein white or red... It would code for different proteins, pink or white or red.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Alleles differ in nucleotide sequence affecting the protein which gives trait variation, dominant and recessive relationships are explained by the interaction of proteins produced</td>
<td>hypothetical: “The alleles have small changes in DNA that code for a red or white pigment protein. In a pink flower, both red pigment proteins and white pigment proteins are made. Since they both get expressed and show, the flower looks pink.”</td>
</tr>
</tbody>
</table>
Table 3.7

*Codes Developed for Construct G1*

**Big Idea**

The genetic information determines how we look and function (phenotype) and such variation in the DNA can serve as a way to identify individuals and species.

**Questions**

1. Check the box next to the statement you think best describes why these organisms look different (fruit fly, human boy, human girl).

2. Which of the following statements do you think best explains why the flowering plants look different?

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No explanation</td>
<td>1025 (pre): “IDK”</td>
</tr>
<tr>
<td>0+</td>
<td>0.33</td>
<td>Organisms have different traits or functions</td>
<td>1056 (pre): “The plants have some alikes and some unalikes.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>Different organisms have different genetic information</td>
<td>1013 (pre): “Roses and daisies have different DNA, thus, they look different.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Different organisms have different genetic information even within a species</td>
<td>1026 (pre): “All flowers are different so their DNA must be different.”</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>The sex chromosomes X and Y vary in boys versus girls or chromosomes vary among plants</td>
<td>2010 (pre): “Plants have some of the same and different chromosomes.”</td>
</tr>
<tr>
<td>2+</td>
<td>2.33</td>
<td>The proteins are different inside of different organisms</td>
<td>1037 (post): “The DNA codes for different proteins which cause different things in the plants.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-</td>
<td>2.67</td>
<td>Some DNA varies between organisms of the same species and some does not</td>
<td>1017 (post): “Their DNA is very similar but they also have some small differences that make them look different.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Some DNA varies between species and some does not (we share some genes with other species)</td>
<td>1026 (post): “Plants have to have the same DNA to photosynthesize but have different DNA to look different in color and shape.”</td>
</tr>
</tbody>
</table>
Changes to the genetic information can cause changes in how we look and function (phenotype) and such variation can lead to the evolution of species over time.

Question

1. Which scientist do you think best explained what would happen to the plants if they survived being “fertilized” with the pesticide?

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No explanation</td>
<td>1038 (pre): “I didn’t really know.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>Different species of organisms look and function differently</td>
<td>1053 (pre): “The flowers would look and function different.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Organisms look and function differently even within a species because they are different organisms</td>
<td>1026 (post): “Every plant has different DNA which means when mutated the plant would react differently than other plants.”</td>
</tr>
<tr>
<td>1+</td>
<td>1.33</td>
<td>Changes to an organism are either always bad or always good</td>
<td>1031 (post): “The plant would change. Wouldn’t the pesticide mutate the plant and cause it not to grow.”</td>
</tr>
<tr>
<td>2-</td>
<td>1.67</td>
<td>Changes could be beneficial, harmful, or neutral to an organism</td>
<td>1006 (post): “Each plant would react differently because the mutations would cause the plants to either grow more or less.”</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>DNA changes could be beneficial, harmful, or neutral to an organism, these changes result in changes to the protein structure and function</td>
<td>1041 (post): “It could be a good change or a bad change. And it also changes the structure of the, the structure and function of the proteins inside. Which in the end could cause physical changes, which I think is right because that’s what happens.”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>DNA changes lead to increased genetic variation over time and evolution of a species</td>
<td>1048 (post): “When things are mutated in the DNA... it is possible that you could pass it on, that they could pass on the mutated DNA, which would cause genetic variation... If the mutation caused the plant to make more seeds ten the mutated plant would produce more and it could pass on its mutation.”</td>
</tr>
<tr>
<td>3+</td>
<td>3.33</td>
<td>If changes were beneficial, the population would shift to look more like the mutated organisms over time (natural selection)</td>
<td>1019 (mid): “If the mutation is good then it would pass on the mutations more. So that the plants in the future will have that as well.”</td>
</tr>
</tbody>
</table>
### Table 3.9

*Codes Developed for Construct H*

#### Big Idea

Environmental factors can interact with our genetic information.

#### Questions

1. Use the statements below to create an explanation for why Fred developed lung cancer and his twin brother did not.

2. Check the box next to the statement you think best explains what would happen to the plants after they were exposed to the pesticide.

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>No explanation</td>
<td>1019 (pre): “Not quite sure.”</td>
</tr>
<tr>
<td>1-</td>
<td>0.67</td>
<td>The environment has no effect on traits or cells or DNA</td>
<td>1050 (pre): “I don’t think it would do anything bad because we test stuff on animals and they might not have the same reaction as us. And a plant’s like really different than a human... I don’t think that it would necessarily like corrupt the plant or something just because it will put cancer on us.”</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>The environment can affect traits and functions</td>
<td>1027 (post): “I’ve seen it happen before myself. Cancer causing pesticides do deform plants. That’s why DDT was banned in the US in the 1990’s.”</td>
</tr>
<tr>
<td>1+</td>
<td>1.33</td>
<td>The environment can affect cells or organs or tissues</td>
<td>1048 (pre): “Cancer is a mutation in human cells so the pesticides could mutate the plant cells as well.”</td>
</tr>
<tr>
<td>2-</td>
<td>1.67</td>
<td>The environment can change or mutate things inside of the cell such as DNA or genes (general statement)</td>
<td>1034 (post): “The DNA would not stay the same if influenced by an outside source (ex. pesticide)”</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
<th>Response Example (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>The environment can change proteins (type and amount), which influence cell function</td>
<td>1053 (mid): “The pesticide would mutate the proteins... causing them to not function properly. Which is what happens in cancer. Something happens to a protein and it doesn’t function right.”</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>The environment can mutate genes which change proteins or alter gene expression of proteins</td>
<td>1041 (post): “So basically when a protein’s under-expressed, it’s not doing its job. And when a protein is over-expressed, it’s doing too much, so like we learned in the cancer unit, some cells will have a protein that, say, don’t let other cells be next to me. So, and cancer’s case, the way tumors grow, that cell’s under-expressed because they just kinda cluster together. And that’s because the DNA is coding bad proteins.”</td>
</tr>
</tbody>
</table>
genetics content in each of the constructs was also mapped to the learning progression using middle interviews (if available) and post-test and post interviews. The Duncan et al. (2009) learning progression was chosen for data analysis over the Roseman et al. (2006) progression because the progression is broken down into eight “Big Ideas” and also describes a learning performance for three different levels for each of the “Big Ideas.” Therefore, student learning can be mapped to discrete levels for each of the eight “Big Ideas” and then followed over time.

The written assessments were coded and analyzed according to rubrics derived from the Duncan et al. (2009) molecular genetics learning progression (Tables 3.1-3.9). Student responses were scored by their justification for choosing an answer on the multiple choice portion of the assessment. If students chose the level 3 multiple choice response but were not able to justify their answer, they were given a score of 0 because the response is the same as a guess. Each “Big Idea” construct has at least one question, so an initial level for each “Big Idea” construct for each student was determined. A final level for each “Big Idea” construct for each student was also determined to show how students progressed through each construct over the course of the molecular genetics instructional period. The mapping illustrated student learning in three different contexts mapped to the Duncan et al. (2009) progression. While not a comparison study, the results documented how different learning contexts influence achievement in molecular genetics related to the learning performances outlined in the progression.

The interviews were transcribed verbatim and were coded and analyzed according to rubrics derived from the Duncan et al. (2009) molecular genetics learning progression (Tables 3.1-3.9). The interview questions refer to the written assessments, so the
interview data correlates with the written assessment data. Each “Big Idea” construct had one interview question (construct G had two questions), so the interviews provide further evidence of the level of the students for each of the “Big Idea” constructs.

Classroom videos were viewed and transcribed when necessary. For the majority of the videos, this was simply a short narrative of the day’s lesson. For a few others, this included transcribing clips of small group work. In the interviews, students were asked to refer to specific things from the instruction that helped them better understand the concepts in molecular genetics. The instances that the students recalled, if they occurred during a video taped lesson, were reviewed and analyzed for evidence of student learning and how the intervention units or classroom instruction helped the students learn the content.

Student written artifacts were also viewed when necessary to help correlate the instructional instances that students recalled with their ideas of molecular genetics concepts. The four sources of data were used for triangulation to determine the individual students’ levels for each of the “Big Idea” constructs in the Duncan et al. (2009) molecular genetics learning progression.

**Rubric to Learning Progression Revision and Refinement**

The previously described rubrics based on the original Duncan et al. (2009) learning progression (Tables 3.1-3.9) were used as a foundation to revise and refine the Duncan progression. While a detailed analysis of the revision and refinement of each construct will be described in Chapter 3, Empirically Testing and Revising the Molecular Genetics Learning Progressions, the heuristics used to add, remove, and split levels were consistent with the heuristics for coordinating empirical data and learning progressions.
outlined by Shea & Duncan (2013) and described in Chapter 1. The different levels on the revised and refined constructs were given corresponding whole number numerical codes (0-5 or 0-6) that allowed students’ responses to be represented by numbers that are ranked and ordinal (Tables 4.1-4.10). The codes are considered ranked because a more sophisticated response corresponds to a higher number than a less sophisticated response. The numbers are considered ordinal as opposed to interval because although the numbers are rank ordered, they are still considered to be categorical because they are not necessarily evenly spaced. For example, in some constructs there could be a large conceptual jump between level 3 and level 4 while in others there could be a smaller conceptual jump between levels 3 and 4. Additionally, it is impossible to quantify the “amount” of conceptual leap between ideas to give even spacing between the different codes.

**Statistical Analyses**

The whole number 0-5 or 0-6 codes described in detail in Chapter 3, Tables 4.1-4.10, were used for all statistical analyses. As described in the text and tables, different levels were given a corresponding numerical code allowing the students’ responses to be represented by numbers that are ranked and ordinal. The number of statistical tests that can be done on ordinal data are limited (Gravetter & Wallnau, 2009), so despite the fact that t-tests, ANOVAs, and ANCOVAs are statistically not recommended for ordinal data, many researchers across all fields including science education (Elkund et al., 2007; Freidenreich et al., 2011; Gelbart, Brill & Yarden, 2009; Gobert & Clement, 1999; Marbach-Ad et al., 2008; Rotbain et al., 2006; Rotbain et al., 2008) use t-tests, ANOVAs and ANCOVAs to determine significance of data.
Researchers do this because obtaining true interval data is extremely difficult, most data collected is actually ordinal in nature, and given the appropriate parametric assumptions, the $t$- and $F$-statistics have been found to be meaningful for ordinal data (Davison & Sharma, 1994).

Student data was analyzed using paired $t$-tests and Wilcoxon matched-pairs signed-ranks test for each “Big Idea” construct. The mean score for each “Big Idea” construct was calculated for each student for each type of assessment (interview and/or written) and the significance of the difference in mean scores for the pre- and post-interviews/tests was determined using a two-tailed paired $t$-test and Wilcoxon matched-pairs signed-ranks test. Since many $t$-tests were done using the same data thus leading to an increased chance of type I error (false positives), significance was established using a post-hoc Bonferroni correction which divides the $\alpha$ level (0.05) by the number of tests being performed (9, one per rubric/construct). The correction reduces the chance of type I error (false positives); however it also increases the probability of false negatives (type II error) and reduces the power (Bonferroni, 1936; Miller, 1981). Despite the reduced power and increased chance of false negatives, the correction was chosen because the main concern is reducing false positives (type I error). The Wilcoxon matched-pairs signed-ranks test determines whether the distribution of scores from two correlated samples are significantly different and “gives more weight to a pair which shows a large difference between the two conditions than to a pair which shows a small difference” (Siegel, 1956). Statistical differences was determined by using these two tests.
Reliability

Coding reliability was established using inter-rater reliability. A minimum of 10% of all the responses (both written and interview) for each of the questions were coded separately by two different researchers. Coding scores were compared and discussed and for each of the coding rubrics, inter-rater reliability was greater than 85%.

Trustworthiness and Observer Effects

Since this research study included qualitative data, efforts were made to ensure trustworthiness in this project. There are four main ways to ensure trustworthiness in qualitative research: credibility, transferability, dependability, and confirmability (reviewed in Shenton, 2004). Credibility is concerned with accurately recording and measuring the phenomena being studied. In order to establish credibility, this study used well established research methods that have been used in prior studies: researchers became familiar with the culture of the schools and students prior to data collection by being present the previous school year in both schools and both classrooms where lessons were video taped and students were interviewed; a random sample of students were selected to interview (as previously described in detail); a variety of data sources were used to triangulate student learning (as previously described in detail); tactics to help ensure honesty in students such as allowing students to refuse to participate and establishing rapport with students and ensuring that there were no right or wrong answers and that their teacher will not view the tapes and their class grade will not be affected; iterative questioning of students was used to determine if students’ answers were consistent with their prior answers; the research project itself has been and will be subjected to peer scrutiny, inter-rater reliability has been calculated when coding data,
“raw” student quotes are shown as illustrative examples of coded data (Tables 3.1-3.9); and the findings were related to and compared with previous research findings.

Transferability is concerned with the findings of the study being able to be transferred to different contexts. Although each classroom context in this study is unique and cannot be exactly replicated, sufficient contextual information can be provided so that other researchers and teachers can compare findings of this study with findings in other situations. To help with transferability, detailed information about each classroom context was provided in Table 2; details about number of participants and how they were selected were provided in previous sections; the different types of data collection were outlined in previous sections, and the written assessments and interview questions are provided in Appendices A-B.

Dependability is concerned with similar data being obtained if the research were to be repeated in the same contexts with the same methods and same participants. To help with dependability, detailed information about the research design and its implementation has been provided in previous sections and the written assessments and interview questions have been provided in Appendices A-B.

Confirmability is concerned with the findings being a result of the ideas of the participants and not the preferences of the researcher. To address concerns with confirmability, data was triangulated from different sources and a theoretical framework used to analyze data was previously described to show the lens through which the data were viewed and analyzed. Additionally, reliability was established between two researchers for each of the nine coding schemes, each scheme having >85% inter-rater reliability.
Since this research involves human subjects and interacting with them, there are concerns about observer effects such as the Hawthorne and Pygmalion effect. The Hawthorne effect is concerned with results of an experiment not being due to the treatment or intervention applied but instead due to the fact that the participants know they are being studied and are the subject of an intervention; while the Pygmalion effect is concerned with teachers influencing student performance based on their knowledge of the treatment effect or expectations that the treatment will have a positive effect (reviewed in Draper, 2010). To combat the Hawthorne effect, students in the study were not told which lessons were “intervention lessons” or even if their teacher was using any researcher-made interventions. The only rewards given for participating in the study was a cookie to each student after completion of all interviews and written assessments and a small gift of classroom supplies to each teacher (classroom set of clipboards or dry erase boards) upon completion of the study. It was also not communicated to the students or the teachers prior to receipt that there would be a reward or gift of any kind for participating; there was no external “motivation” to do well to receive a reward. Although it is impossible to determine the goals for each individual participating in this study and how they may have affected performance, attempts were made to minimize observer effects.

To combat the Pygmalion effect, teachers were not told of any learning performance expectations of students with or without the intervention units. Teacher effects are much more difficult to counteract. As reviewed in Draper (2010), a large issue in educational research in general is teacher effects and the effect of different teachers are nearly always bigger than the effect of different treatments. To deal with the teacher
effects, the different classrooms will not be directly compared but student achievement under each of the three contexts will be studied and mapped to the learning performances described in the Duncan et al. (2009) learning progression. There is no way to directly compare the different classrooms due to teacher effects; classroom, school, and district contexts; different students; and different curriculum. There are simply too many variables to complete a direct comparison between each classroom; consequently student achievement in three different contexts will mapped to the learning performances described in the Duncan et al. (2009) progression and will be analyzed and used to revise and refine the progressions; in addition the effects of the intervention units on student achievement in different contexts will be studied. Student achievement in each classroom will not be directly compared as a control group versus experimental group study.

Confidentiality and Risk Assessment

The confidentiality of students and teachers was taken very seriously. Consent forms were provided to the students and the teachers. Students and their parents or guardians had the option to agree or not agree to participate in each of the three types of data collection: classroom videos, interviews, and written assessments. If a parent or guardian refused to allow a student to participate, the decision involved no penalty. Neither the student’s grades, school records, nor the student’s ability to participate in the class was affected. Teachers could withdraw themselves and their classrooms from the study at any time without penalty. Parents or guardians could withdraw their students at any time from the study without penalty. If the parents initially gave permission to allow their student to be video taped and later changed their mind, the parent and/or the student always had the right to ask permission to be outside of the video camera’s view.
Pseudonyms were used for every student and teacher in the data collection documents. The pseudonyms were used on all written work (such as assessments, student classroom work) and all interviews. Student names were removed from digital copies of scanned student work and were replaced with pseudonyms. All records were kept strictly confidential. Each teacher’s name and student’s name has been and will be kept strictly confidential in all publications and presentations.

The presence of other people in a classroom can be stressful, so a close cooperation with the teachers was necessary. Arrangements for videotaping and interviews were made in advance and no videotaping was done without advanced notice. If a teacher did not want certain activities or lessons taped, no data was collected on those activities or lessons. Researchers did not interfere with ongoing classroom activity and assured all the students and teachers that they are there to investigate how students learn molecular genetics, not to evaluate them. Everything was done to ensure that the teacher and students had a positive learning environment during data collection.

If a classroom was videotaped and permission had not been given to videotape a particular student, the student was not specifically videotaped. However, the video recordings did capture the entire classroom; those students who did not give consent to be videotaped could not be fully excluded from the videotaping. It was attempted to situate students who did not express consent to the study out of the view of the camera. Data were not analyzed from those students who did not give consent to the study. Anything these students did or said was not transcribed or used in any analysis and a videotape or image or voice of any non-consented student was not and will not be used in any way.
The students whose parents or guardians have not given permission were able to fully participate in the classroom activities but were not included in transcription or analysis.

**Institutional Review Board Approval**

This study has been approved by the Wright State University IRB board, SC# 4547.
III. Results

Empirically Testing and Revising the Molecular Genetics Learning Progressions

Student understandings of molecular genetics were assessed by both written assessments and interviews in three 10th grade classroom contexts. The assessments and interviews targeted each of the eight “Big Idea” constructs of the Duncan et al. (2009) learning progression, with each construct having at least one written question and one interview question. It was expected that students would fall on the extremely low levels of achievement and learning performances before classroom instruction and progress to the higher levels of achievement as the instructional period progressed, allowing for tracking student progress through the entire progression.

It was hypothesized that students would hold ideas not included in the molecular genetics learning progressions, the empirical data obtained in the study could be used to revise and refine the progressions, and students who completed the activities in the intervention units would achieve higher learning performances in constructs targeted by the units than the students who did not complete the intervention units. Students in all contexts did hold many ideas not defined on the progressions. The empirical data was used to revise and refine the progressions using the general heuristics provided by Shea & Duncan (2013). However, students who completed the activities in the intervention units did not necessarily achieve higher learning performances in constructs targeted by the units than the students who did not complete the intervention units.
The following sections describe student achievement in each of the eight “Big Idea” constructs of the Duncan et al. (2009) learning progression and the specific revisions and refinements for each construct. The revised levels were used for all statistical analyses to determine significance of student learning performance increases and correspond with the percent of students at each level for each of the constructs described in Figures 3-11. Additionally the following sections include combining the Roseman et al. (2006) and Duncan et al. (2009) progressions into one progression incorporating the ideas from both progressions.

**construct A.** This construct focuses on the idea that genetic information is hierarchically organized and was not specifically included in any of the intervention units. In the original Duncan et al. (2009) learning progression, this construct contained three levels: 1-3 (Table 4.1). Upon empirical testing, these levels were found to be valid and represented ideas that many students had. Additionally, several lower and intermediate ideas were found.

In the original progression, the levels contained defined relationships between specific words; for example, level 1 described that genes are in cells and level 2 described that genetic information is found in chromosomes of cells (Table 4.1). When looking at empirical data, some students were able to describe these specific correlations, but many made other correlations between the words. The idea that chromosomes are inside of cells was not found in the original progression. To combat this issue, the level descriptions were changed from defined relationships between words to being able to describe relationships between more words as students progressed through the levels. On the revised progression, the idea that genes are in cells is at the same level as the idea that
Table 4.1

Empirical Revisions and Refinements of Construct A

<table>
<thead>
<tr>
<th>Original LP Duncan et al. (2009)</th>
<th>Revised LP Todd (2013)</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0</td>
<td>Level 0</td>
<td>No correlation between words (DNA, gene, chromosome, nucleotide/base, cell, genome)</td>
</tr>
<tr>
<td>Level 1</td>
<td>Level 1</td>
<td>All incorrect correlations between words</td>
</tr>
<tr>
<td>Level 2</td>
<td>Level 2</td>
<td>Correct correlation between 2 words</td>
</tr>
<tr>
<td>Humans, animals, plants, fungi, and bacteria have genes (genetic information) in their cells</td>
<td>The genetic information is found in the chromosomes of cells. Most sexually reproducing organisms have two sets of chromosomes. All cells of an organism have the same two chromosomal sets (except sex cells)</td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 3</td>
<td>Correct correlation between 3-4 words</td>
</tr>
<tr>
<td>Genes are nucleotide sequences within the DNA molecule. DNA molecules make up chromosomes that make up our genome.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>Level 4</td>
<td>Correct correlation between 5 words</td>
</tr>
<tr>
<td>Level 5</td>
<td>Level 5</td>
<td>Correct correlation between all 6 words</td>
</tr>
</tbody>
</table>

Note. Strikethrough indicates an idea that was removed from the construct. Black arrows indicate similar ideas. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, Journal of Research in Science Teaching, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
chromosomes are in cells. These both describe a correct correlation between two words and would be a level 2 response (Table 4.1).

Prior to instruction, students in each context had a very basic understanding of genetic organization (Figures 2A, 2B). Many students were not able to describe a correlation between any of the words or only described incorrect correlations (Figure 2A, 2B, levels 0-1). These ideas described lower learning performances than were included in the original Duncan et al. (2009) progression. The ideas are directly related to the content and do represent important conceptual shifts in student thinking and are productive stepping stones, so they were included as new levels in the construct (Table 4.1, levels 0-1). Looking at the student responses to a specific question in this construct (Q1: put the following terms in some sort of order or pattern: DNA, gene, chromosome, nucleotide/base, cell, genome) in more depth, approximately half of the students in context 1 and context 3 were unable to describe a correlation (correct or incorrect) between any words (level 0) on the written assessment prior to instruction (Figure 3A). However, 75% of students in context 2 were able to describe a correlation (correct or incorrect) between two or more of the words (levels 1-5) on the pre written assessment. On the pre interviews (Figure 3B), the percent of students in each level were very similar with approximately 90% of students in context 1 and 100% of students in context 2 able to describe a correlation (correct or incorrect) between two or more of the words (levels 1-5). This indicates that students in the three different contexts likely entered the molecular genetics instructional period with similar understandings of genetic organization, but students in context 2 were better able to describe the correlations on the written pre assessment than students in the other contexts.
**Figure 2.** Averages of written assessments and interviews. Averages of student responses for pre and post written assessment questions (A) or pre, middle, and post interview questions (B) for each of the “Big Idea” constructs of the Duncan et al. (2009) learning progression. *p < 0.0055 on pre versus post. Inter-rater reliability > 85% for each construct."
Figure 3. Percent of students at each level of construct A. Percent of students at each level of construct A on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q1: Put the following terms in some sort of order or pattern: DNA, gene, chromosome, nucleotide/base, cell, genome. Inter-rater reliability > 85%.
After instruction, student achievement significantly increased in each context after instruction on both the written assessments (Figure 2A) and interview (Figure 2B). Over 75% of students in contexts 1 and 2 were able to correctly describe a correlation between at least three of the six words (levels 3-5) on the written assessments (Figure 3A). Just less than half of the students in context 3 were able to achieve this level. Several students in the three contexts were able to describe a correct correlation between 5 of the 6 words, most often having difficulty with correctly describing genome (Figure 2A, 2B, level 4). This idea also represents an important conceptual shift in student thinking and is a productive stepping stone, so this level was also added to the construct (Table 4.1, level 4). Students in context 1 were the only students able to describe correct correlations between all six of the words (level 5). Nearly 25% of students were able to do so on both the written assessments (Figure 3A) and interviews (Figure 3B). This indicates that students in context 1 made the largest gains in understanding genetic organization, followed by students in context 2. However, student achievement in each of the three contexts significantly increased after instruction on both written assessments and interviews.

This construct also originally included the idea that humans, animals, plants, fungi, and bacteria have genes or genetic material (Table 4.1, original LP level 1). Although this idea is important, it does not seem to fit well with the rest of the construct which deals with the organization of genetic material and was thus removed from construct A. The idea that all these organisms have genetic material may be better suited to construct G which describes genetic similarities and differences between organisms; this will be discussed in the section about construct G.
**construct B.** This construct focuses on the idea that genes code for protein structure and function. This idea was included in the third intervention unit (“Can we engineer a superhuman?”) which was only taught to students in context 1. Like all of the construct ideas, the concept of genes coding for protein structure and function is normally taught in 10th grade biology; however the intervention unit provided additional inquiry activities to help students achieve the highest level in the construct. In the original Duncan et al. (2009) learning progression, this construct contained three levels: 1-3 (Table 4.2). Shea & Duncan (2013) refined this construct as an example of how to use empirical data to refine learning progressions. The revisions lead to an additional 4 levels on this construct (Table 4.2). Upon empirical testing in this study, the revised levels of this construct were found to be valid; the revised levels represented ideas that a large majority of students had.

The additional four levels of the construct added by Shea & Duncan (2013) were very helpful for categorizing student understandings of the function of genes. Based on the data obtained in this study, a new level (Table 4.2, level 5) was added to the construct, including the idea that genes code only for proteins which are made of amino acids. This idea represents a shift in student thinking and is a productive stepping stone, so it should be included in the progression. A level 4 understanding is that genes code for entities inside of cells (Table 4.2). Students may explain that genes code for molecules or amino acids or proteins, but they do not understand that genes code only for proteins. A level 6 response is an understanding that genes are translated into specific amino acid sequences that make up proteins (Table 4.2). In Shea & Duncan’s (2013) revisions of the progression, the idea that genes code only for proteins was missing. The data in this
### Table 4.2

**Empirical Revisions and Refinements of Construct B**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Description</td>
<td>Level Description</td>
<td>Level Description</td>
</tr>
<tr>
<td>Level 0</td>
<td>No knowledge of genes.</td>
<td>Level 0</td>
</tr>
<tr>
<td>Level 1</td>
<td>Genes are non-informational in nature. They are passive particles associated with traits.</td>
<td>Level 1</td>
</tr>
<tr>
<td>Level 2</td>
<td>Genes are non-informational in nature. They are active particles associated with traits.</td>
<td>Level 2</td>
</tr>
<tr>
<td>Level 3</td>
<td>Genes are active instructions that “tell” proteins, the cell, or the body to carry out specific functions.</td>
<td>Level 3</td>
</tr>
<tr>
<td>Level 1</td>
<td>Genes are instructions for how organisms grow, develop, and function</td>
<td>Level 4</td>
</tr>
<tr>
<td>Level 4</td>
<td>Genes have information about biological entities and function at multiple organization levels.</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
<th>Revised LP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2009)</td>
<td>Level Description</td>
<td>Revised LP</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>Level 5</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>Level 6</td>
</tr>
<tr>
<td></td>
<td>Genes are instructions for molecules (many of which are proteins) that carry out functions within an organism. All organisms use the same genetic language for their instructions</td>
<td>Genes are instructions for molecules (many of which are proteins) that carry out functions within the organism. All organisms use the same genetic language for their instructions.</td>
</tr>
<tr>
<td></td>
<td>The genetic code is translated into a sequence of amino acids that make up the protein. Almost all organisms use the same genetic code</td>
<td>The genetic code is translated into a sequence of amino acids that make up the protein. Almost all organisms use the same genetic code.</td>
</tr>
</tbody>
</table>

**Note.** Dashed arrows indicate that two ideas were combined. Black arrows indicate similar ideas. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc. and “From Theory to Data: The Process of Refining Learning Progressions,” by N. A. Shea, and R. G. Duncan, 2013, *The Journal of the Learning Sciences*, 22, pp. 15. Copyright 2013 by Taylor & Francis Group, LLC.
study indicate that many students do hold this idea after instruction (Figure 4, level 5) and that understanding that genes code only for proteins is a productive step to understanding how genes are translated into the specific amino acid sequences of proteins.

Overall prior to instruction, students in each context had an extremely basic understanding of what genes do, but student achievement significantly increased after instruction in contexts 1 and 3 on the written assessments (Figure 2A) and both context 1 and 2 on the interviews (Figure 2B). Looking at the student responses to a specific question in this construct (Q3: check the box next to the statement you think best explains how DNA is involved in muscle function) in more depth, approximately 60% of students in context 1, 75% of students in context 2, and 85% of students in context 3 were not able to describe any functions of genes (level 0) on the written pre assessment (Figure 4A). More students in context 1 were able to explain that DNA contains information (levels 2-6) on both the written assessments (30%, Figure 4A) and interviews (55%, Figure 4B) than students in context 2 (18% and 40%, respectively) and context 3 (5%) indicating that students in context 1 may have entered the year with slightly more knowledge of the function of genes than the students in the other two contexts, although not substantially.

After instruction, nearly 50% of students in context 1 were able to describe that genes are informational in nature on their written post assessments (Figure 4A). Eighteen percent of students in context 2 and 30% of students in context 3 were able to describe this on their written post assessments (Figure 4A). Thirty-five percent of students in context 1 and 5% of students in context 3 were able to describe in their written post assessments that genes code for entities inside of the cell (such as proteins or molecules) that carry out functions (Figure 4A, levels 4-6).
Figure 4. Percent of students at each level of construct B. Percent of students at each level of construct B on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q3: Check the box next to the statement you think best explains how DNA is involved in muscle function. Inter-rater reliability > 85%.
A much more dramatic change was seen in the student interviews. Over 95% of students in context 1 were able to describe that genes code for entities inside the cell that carry out functions (Figure 4B, levels 4-6) with over 25% of students able to describe how DNA is translated into amino acids that make up a protein (level 6). In context 2, just over 25% of students were able to describe that genes code for things inside of the cell that carry out functions (levels 4-6) and just under 5% of students described how genes were translated into proteins (level 6). These data document that a large portion of the students in context 1 made a very dramatic shift from a basic understanding of genes prior to instruction to a much more complex idea that genes code for biological entities inside of the cell after instruction. Students in context 2 also increased their understandings of genes a bit after instruction but were unable to describe their increased understandings in writing. Students in context 3 showed, in writing, a better understanding of genes than context 2 (Figure 4A), but only 5% of students explained that genes code for biological entities inside cells (levels 4-6). The relatively poor student performances in this construct after instruction in two of the three contexts is further evidence that the idea, and thus the molecular model of genetics, is a very challenging concept for students to learn.

During the revision of their progression, Shea & Duncan (2013) added levels describing the understandings that genes are non-informational in nature and passive particles, and that genes are non-informational in nature and active particles (Table 4.2, Shea & Duncan revised levels 1-2) which comes from the work of Venville & Treagust (1998). Both understandings include the idea that genes are non-informational in nature. The very fine distinction between the two levels is that a student with a passive particle
understanding will explain that “genes are your traits” while a student with an active particle understanding will explain that “genes make your traits.” Neither explanation includes any reference to genes containing information, thus both are non-informational understandings of genes. Although Shea & Duncan (2013) did find students in these levels, Duncan herself later indicated that she was not sure how useful it is to tease students with an active versus passive view apart from one another because both understandings are non-informational in nature (Ravit Golan Duncan, personal communication).

Data from this study indicated that very few students held a non-informational view of genes before or after instruction (Figure 4, level 1) and it was extremely likely that no data was lost by combining the non-informational passive and non-informational active understandings into one level. Therefore, the suggested revisions and refinements of construct B include combining the non-informational levels (Table 4.2, Shea & Duncan revised levels 1-2) into a singular non-informational level. The important lower stepping stones for this construct are going from no knowledge of genes (Table 4.2, Todd revised level 0) to an idea that genes are passed down (Table 4.2, Todd revised level 1) to an idea that genes contain information (Table 4.2, Todd revised level 2). Once students understand that genes contain information, they can then learn what that information codes for (Table 4.2, Todd revised levels 3-5) and how genes are translated into a specific order of amino acids to make up proteins (Table 4.2, Todd revised level 6).

**construct C.** This construct focuses on the idea that proteins have a central role in the functioning of organisms and are the mechanism that connects genes and traits. This idea was included in all three of the intervention units since all the units described
roles of proteins and gave examples of protein functions. In the original Duncan et al. 
(2009) learning progression, this construct contained three levels: 1-3 (Table 4.3). Shea 
& Duncan (2013) also refined this construct as an example of how to use empirical data 
to refine learning progressions. Their revisions lead to an additional 3 levels on this 
construct (Table 4.3). They also refocused the construct on the singular idea that proteins 
have a central role in the functioning of organisms, omitting the ideas included in the 
original progression that dealt with how changes to genes affect traits (Table 4.3, 
strikethrough text on original LP). When assessments were created for this study, the 
assessments for this construct targeted the idea that proteins connect genes and traits. 
Due to this construct having two ideas and the Shea & Duncan (2013) revisions omitting 
one of the ideas, the suggested revisions and refinements in this study include dividing 
this construct into two constructs: C1 (proteins have a central role in the functioning of 
organisms, Table 4.3) and C2 (proteins are the mechanism that connect genes and traits, 
Table 4.4).

Although the idea that proteins have a central role in the functioning of organisms 
was not specifically addressed in this study, when talking with students about proteins 
and hearing them explain what proteins are and what they do both during interviews and 
in classroom discourse, the revised levels of C1 proposed by Shea & Duncan (2013) are 
ideas that students have and likely represent valid levels of understandings students have 
in the three contexts in this study (Table 4.3). Assessment items specifically addressing 
this idea would need to be created to further validate Shea & Duncan’s (2013) findings, 
but it is likely that the revised levels of this construct will be found valid. The only 
modification proposed to construct C1 is the removal of the idea that changes to the gene
Table 4.3

Empirical Revisions and Refinements of Construct C1

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2009) Level 1</td>
<td>Shea &amp; Duncan (2013) Level 1</td>
<td>Cells have to carry out many essential functions to live. Within cells organelles do specific functions. The structure of cells, tissues, and organs determines their function. Our body has multiple levels of organization and changes at one level may affect another</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cells are one of the basic levels of organization in the body. They can perform specific functions.</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>Proteins are like little machines that do the work of the cell. Proteins have shapes and properties that afford their functions. There are different types of proteins (enzymes, receptors, etc.) Changes to genes can result in changes to proteins, which can affect the structures and functions in the organism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proteins are good for you and provide positive health benefits. Without them, general health declines.</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>Protein function is dependent upon their structure. No examples given of how protein function (or a lack of function) contributes to the genetic phenomena.</td>
</tr>
<tr>
<td></td>
<td>Level 4</td>
<td>Protein function is dependent upon protein structure. Specific examples given of how protein function (or a lack of function) contributes to the genetic phenomena.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP Duncan et al. (2009)</th>
<th>Level Description</th>
<th>Revised LP Shea &amp; Duncan (2013)</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 3</td>
<td>Proteins have particular three-dimensional shape determined by their amino acid sequence. Proteins have many different kinds of functions that depend on their specific properties. There are different types of genetic mutations that can affect the structure and thus function of proteins and ultimately the traits</td>
<td>Level 5</td>
<td>The structure of proteins is dependent upon the properties of amino acids that make up the protein. Changes in a gene can result in changes to specific amino acids, resulting in protein structure/function changes or loss of function.</td>
</tr>
</tbody>
</table>

*Note.* Strikethrough indicates an idea that was removed from the construct by Shea & Duncan. Black arrows indicate similar ideas. Grey background indicates an idea that was added to the construct by Shea & Duncan. Italicized strikethrough indicates an idea removed from the construct as a result of this study.

### Table 4.4

**Empirical Revisions and Refinements of Construct C2**

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Level Description</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2009)</td>
<td><strong>Level 0</strong> Genes and traits are not connected</td>
<td><strong>Level 0</strong></td>
<td>Genes and traits are not connected</td>
</tr>
<tr>
<td>Level 1</td>
<td>Cells have to carry out many essential functions to live. Within cells, organelles do specific functions. The structure of cells, tissues, and organs determines their function. Our body has multiple levels of organization and changes at one level may affect another</td>
<td><strong>Level 1</strong></td>
<td>Changes in genes can change instructions or traits</td>
</tr>
<tr>
<td>Level 2</td>
<td>Proteins are like little machines that do the work of the cell. Proteins have shapes and properties that afford their functions. There are different types of proteins (enzymes, receptors, etc.) Changes to genes can result in changes to proteins, which can affect the structures and functions in the organism</td>
<td><strong>Level 2</strong></td>
<td>Changes in genes can change cells, genes tell cells what to do</td>
</tr>
<tr>
<td><strong>Level 3</strong></td>
<td>Changes to a gene change proteins (no mention of protein-trait connection)</td>
<td><strong>Level 3</strong></td>
<td>Changes to a gene change proteins (no mention of protein-trait connection)</td>
</tr>
<tr>
<td><strong>Level 4</strong></td>
<td>Changes to a gene change proteins to change traits (no explanation of how)</td>
<td><strong>Level 4</strong></td>
<td>Changes to a gene change proteins to change traits (no explanation of how)</td>
</tr>
<tr>
<td><strong>Level 5</strong></td>
<td>Changes to the gene change amino acids that make up proteins</td>
<td><strong>Level 5</strong></td>
<td>Changes to the gene change amino acids that make up proteins</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP</th>
<th>Level Description</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2009)</td>
<td><strong>Level 3</strong> Proteins have particular three-dimensional shape determined by their amino acid sequence. Proteins have many different kinds of functions that depend on their specific properties. There are different types of genetic mutations that can affect the structure and thus function of proteins and ultimately the traits</td>
<td><strong>Level 6</strong> Changes to the gene can change the function of a protein and ultimately the traits</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Strikethrough indicates an idea that was removed from the construct. Black arrows indicate similar ideas. Dashed arrows indicate an idea was split into levels. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
change proteins in the highest level of this construct (Table 4.3, level 5, italicized strike through text). The impact of changes to genes has been moved to construct C2 and will now be discussed in further detail.

Upon empirical testing, the original levels and ideas of this construct dealing with the connection between genes and traits were found to be valid and represented ideas that many students had. Additionally, lower and intermediate ideas that represent important conceptual shifts and productive stepping stones were found as well. Prior to instruction, students in all contexts had essentially no knowledge of how genes and traits are connected on both written assessments (Figure 2A) and in interviews (Figure 2B). Looking at the student responses to a specific question in this construct (Q4: which student do you think best explained how the change in the gene leads to the physical effects seen with muscular dystrophy?) in more depth, roughly 90% of students in each of the three contexts were not able to explain how genes and traits are connected prior to instruction on both written assessments (Figure 5A, level 0) and in interviews (Figure 5B, level 0). Since a large number of students had no knowledge of how genes and traits were connected, an idea not found in the original Duncan et al. (2009) progression (Figure 5, level 0), this idea is included in the modified construct C2 as level 0 (Table 4.4). The construct deals exclusively with the molecular model of genetics, specifically with the role of proteins; the very poor performance of students across the three contexts on pre assessments and pre interviews is further evidence that students have extremely limited understandings of the molecular model and the role of proteins entering the 10th grade.
Figure 5. Percent of students at each level of construct C. Percent of students at each level of construct C on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q4: Which student do you think best explained how the change in the gene leads to the physical effects seen with muscular dystrophy? Inter-rater reliability > 85%.
Student achievement on written assessments significantly increased after instruction only in context 1 (Figure 2A). In interviews, student achievement significantly increased in both contexts 1 and 2 (Figure 2B). After instruction, approximately 31% of students in context 1, 7% of students in context 2, and 10% of students in context 3 were able to describe on the written assessments that a change in a gene would change a protein which would change traits (Figure 5A, levels 4-6). A more dramatic shift is seen on the post interviews with over 75% of students in context 1 and 18% of students in context 2 able to relate how changes to the gene change proteins to change traits (Figure 5B, levels 4-6). The large gains indicated in context 1 on both the written assessments and interviews indicate a significant shift in student understandings of how genes, proteins, and traits are connected by students in this context.

Conversely, on the written assessments, students in contexts 2 and 3 showed no significant increases in understandings after instruction. The students in context 2 were better able to verbally describe in interviews how genes and traits are related than on their written assessments; context 2 and 3 students did show a significant increase in understandings after instruction on the interviews, but more than 75% of students were still not able to describe how changes to a gene change proteins (Figure 5B, levels 0-3). The high percentage of students still holding very basic ideas of genes after instruction indicates that this is a difficult concept for students to understand. However, given that over 75% of students in context 1 were able to describe how changes to the genes change proteins to change traits after instruction during post interviews (Figure 5B, levels 4-6), it is not unreasonable to expect that 10th grade students are able to achieve this level of
understanding; it does suggest that the methods of instruction and instructional materials must be targeted to upper level ideas.

Several students, mainly students in context 1 after instruction, explained how changes to genes change proteins (but were unable to explain that proteins and traits are connected). The idea represents an important conceptual shift in student thinking and is a productive stepping stone for students since it is a shift in thinking about genes coding for entities at the cellular level or larger (a lower level response) to thinking that genes code for entities at the subcellular level (a higher level response). Although students are not able to explain that proteins play a role in visible traits, students are able to make a connection that genes code for proteins, hence changes to the gene would change a protein. This idea is included in the revised construct C2 as a level 3 response (Table 4.4). Figure 5 reflects that a few students had this understanding (level 3), mostly students in context 1 after instruction.

The idea that changes to the gene change amino acids is also another important conceptual shift that was missing on the original construct. In getting to this understanding, students are able to further explain that genes code for the amino acids that make up a protein; however students are lacking the ability to describe how the changed function of that protein produces the visible trait. Many students were able to describe the relationship between genes and traits to this level (Figure 5, level 5) after instruction, indicating that this is an idea that students have and is a productive stepping stone very close to being able to describe how the altered function of the protein is involved in the visible trait. This idea is included on the revisions and refinements to construct C2 as level 5 (Table 4.4).
Level 1 of the original construct centered on the idea that the body has multiple levels of organization and that changes at one level may affect another. Students had many ideas that were encompassed by this general idea, so this idea was divided into two distinct ideas: changes in genes change instructions or traits (Table 4.4, level 1), and changes in genes change cells (Table 4.4, level 2). Splitting this level provides valuable data to understand if students think of genes as coding for traits (level 1) or for entities at the cellular level (level 2). This distinction is important because traits are visible to the naked eye while cells are not; breaking this idea into two levels provides insight about when students progress from thinking about genes as coding for visible entities such as traits to thinking about genes as coding for non-visible entities such as cells and proteins.

**construct D.** This construct focuses on the idea that all somatic cells have the same DNA but express different genes and proteins. This idea was included in the first intervention unit (“How do cells become cancerous?”), which was taught to both context 1 and 2 students in its entirety. In the original Duncan *et al.* (2009) learning progression, this construct contained three levels: 1-3 (Table 4.5). Upon empirical testing in this study, the levels of this construct were found to be valid and represented ideas that many students had. Additionally, several lower and intermediate ideas were found.

Prior to instruction, students in all contexts had a very basic understanding of how and why cells are different on both written assessments (Figure 2A) and in interviews (Figure 2B). Reviewing the student responses to a specific question in this construct (Q7: which student do you think best explained why the cells [skin and nerve] looked different?) in more depth, over 75% of students in all contexts were unable to describe why cells are different on the pre written assessments (Figure 6A, level 0). This idea was
### Table 4.5

**Empirical Revisions and Refinements of Construct D**

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Level Description</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan <em>et al.</em> (2009)</td>
<td>Level Description: Different cells have some common and some different structures and functions</td>
<td>Todd (2013)</td>
<td>Level Description: No explanation about why cells are different</td>
</tr>
<tr>
<td></td>
<td>Level 0</td>
<td></td>
<td>Level 1</td>
</tr>
<tr>
<td></td>
<td>Level 1</td>
<td></td>
<td>Level 2</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td></td>
<td>Level 3</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td></td>
<td>Level 4</td>
</tr>
<tr>
<td></td>
<td>Level 5</td>
<td></td>
<td>Level 5</td>
</tr>
<tr>
<td></td>
<td>(continued)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original LP</td>
<td>Level Description</td>
<td>Revised LP</td>
<td>Level Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Duncan et al. (2009)</td>
<td>Level 3</td>
<td>All cells have the same genetic content, but what genes are used by the cell (expressed) is regulated</td>
<td>Todd (2013)</td>
</tr>
</tbody>
</table>

Note. Black arrows indicate similar ideas. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
Figure 6. Percent of students at each level of construct D. Percent of students at each level of construct D on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q7: Which student do you think best explained why the cells [skin and nerve] looked different. Inter-rater reliability > 85%.
missing from the original progression and was added as the lower anchor of this construct (Table 4.5, level 0).

Students were better able to describe why cells look different during the pre interviews. Over 60% of students in context 1 and 45% of students in context 2 were at least able to describe that cells look different because they have different functions (Figure 6B, levels 2-6). Many students had a basic understanding of DNA as containing information and explained that the DNA “tells” the cells to be different (Figure 6, level 3). This idea was also not included in the original progression, but represents an important conceptual shift in student understandings from cells being different due to their functions or responsibilities (Table 4.5, level 2) to the idea that something inside the cells themselves tells them or makes them different (Table 4.5, level 3). Because this was an important conceptual shift, it was included on the revised progression as a level 3 response (Table 4.5).

Only 13% in context 1 and 8% of students in context 2 were able to explain that different cells contain different proteins (Figure 6B, levels 4-6) during the pre interviews. This result indicates that roughly half of students enter their 10th grade year understanding that cells are structured for the function they perform, but very few students understand the role proteins play in functions critical to each type of cell.

Student achievement on written assessments significantly increased after instruction in contexts 1 and 3 (Figure 2A). In interviews, student achievement significantly increased in both contexts 1 and 2 (Figure 2B). After instruction, over 60% of students in context 1 were able to describe on written assessments that the differences in cells related to their different functions (Figure 6A, levels 2-6). Only a little over 30%
of students in contexts 2 and 3 were able to describe this idea on their written post assessments (Figure 6A, levels 2-6).

A small number of students in context 2 described that cells look different because they are in different places in the body (Figure 6, level 1). Although it was only a small number of students in context 2 that held this idea, it could be a productive stepping stone to students understanding structure-function relationships because different parts of the body have different functions. Additionally, since this study was done with 10th grade students (the upper limit of the progression), this naive idea may appear more frequently in younger students. Since this idea did appear in this study and it could be a useful stepping stone to learning about structure-function relationships, especially for younger students, the idea was included in the revised progression as a level 1 understanding (Table 4.5).

Over the course of the molecular genetics instructional period, students learned more about gene expression and held various ideas about gene expression. Twenty-five percent of context 1 students were able to further explain on the written post assessments that the cellular differences were due to different proteins inside the cells (Figure 6A, levels 4-6); not quite 4% of students in context 2 and 8% of students in context 3 were able to achieve this level of understanding on the written assessments (Figure 6A, levels 4-6). The original progression contained the ideas that cells contain different repertoires of proteins and that cells have the same DNA but express different proteins (Table 4.5, original levels 2-3). A large number of students in the three contexts did express these ideas, especially after instruction (Figure 6, levels 4 & 6); but some students explained that cells contain different proteins and the same DNA but were unable to elaborate on a
mechanism by which this occurs. This was an intermediate idea between the two original levels but represents an important conceptual shift in student thinking. Students who explain this understanding know that cells contain different proteins for their functions, yet all have the same DNA; the students fail to understand gene expression. Since this idea is an important conceptual shift in student thinking, it was also included on the revised progression as a level 5 response (Table 4.5).

Students across both contexts 1 and 2 were able to describe more sophisticated understandings verbally during the post interviews. Over 92% of students in context 1 were able to explain that different cells have different repertoires of proteins (Figure 6B, levels 4-6); just over 25% of students were able to achieve this level of understanding in context 2 (Figure 6B, levels 4-6). Additionally, only students in context 1 were able to achieve the highest level in this construct and explain that specialized cells contain the same DNA but that expression of the proteins is specific to that type of cell. Just over 4% of students explained this on their written assessments (Figure 6A, level 6) and over 42% of students did so on their post interviews (Figure 6B, level 6). The large gains seen in context 1 students indicate that students can understand this complex idea. However, the much more moderate gains seen in the written assessments of students in context 3 and the interviews of context 2 students indicate that this is another difficult concept for students to understand.

**construct E.** This construct focuses on the idea that organisms reproduce by transferring their genetic information to the next generation. This idea was not specifically included in any intervention units, but is included in typical course instruction of 10th grade biology. In the original Duncan et al. (2009) learning
progression, this construct contained three levels: 1-3 (Table 4.6). Upon empirical testing in this study, the levels of this construct were found to be valid and represented ideas that many students had. As with the other constructs, lower and intermediate ideas were also found.

Prior to instruction, students in all contexts had very simple understandings of how genetic information is passed on to future offspring in written assessments (Figure 2A) and in interviews (Figure 2B). Reviewing the student responses to a specific question in this construct (Q15: check the box next to the statement you think best explains what a baby bunny would look like) in more depth, more than 60% of students in all contexts were unable to reason that DNA or traits are passed down from parents to offspring on the written pre assessments (Figure 7A, level 0). While some students were unable to describe any knowledge of genetic transfer between generations (Figure 7A, level 0), some described that traits (and not genetic information) were passed between parents and offspring (Figure 7A, level 1). These ideas were not included in the original progression, yet represent the lower anchor of this construct and a productive stepping stone on this construct. No knowledge of genetic transfer between generations was included in the modified progression as the lower anchor (Table 4.6, level 0) and the idea that traits were passed down was included as a level 1 response (Table 4.6).

Students were better able to explain their understandings on the pre interviews with approximately 85% of students in both contexts at least describing that DNA or traits are passed down to offspring (Figure 7B, levels 1-5). Prior to instruction, only a small percentage of students in context 1 were able to correctly solve a simple dominant/recessive Punnett square problem (less than 2% on the written assessments, Figure 7A,
Table 4.6

*Empirical Revisions and Refinements of Construct E*

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Level Description</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan <em>et al.</em> (2009)</td>
<td>Original LP</td>
<td>Revised LP</td>
<td>Level Description</td>
</tr>
<tr>
<td>Level 1</td>
<td>All organisms reproduce and transfer their genetic information to their offspring. Cells divide to make new cells each with all the genetic information. In larger organisms each parent contributes half the genetic information to the new generation</td>
<td>Todd (2013)</td>
<td>Level 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Traits or DNA are passed down, organisms can only get traits of their parents</td>
</tr>
<tr>
<td>Level 2</td>
<td>Before cells divide the chromosome sets are duplicated and then two new cells are formed each with two chromosomal sets. In sexually reproducing organisms chromosome sets are randomly assorted into gametes through the process of meiosis (one full set in each sex cell). This process creates sex cells that have only one set of chromosomes</td>
<td>Todd (2013)</td>
<td>Level 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNA is passed down; in sexually reproducing organisms each parent contributes half of the genetic information to the new generation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Level 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Alleles are randomly assorted, each sex cell contains only one allele and these sort independent of chromosomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Level 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In sexually reproducing organisms, cells go through meiosis to create gametes; chromosomes are randomly assorted and each sex cell contains only one set of chromosomes</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level Description</strong></td>
<td><strong>Level Description</strong></td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 5</td>
</tr>
<tr>
<td>DNA replication is tightly regulated to prevent errors. During the process of meiosis chromosomes can swap sections and create new combinations of gene versions on a given chromosome. This creates more genetic variation</td>
<td>Chromosomes can swap sections and recombine during meiosis, creating new combinations of genes on a chromosome increasing genetic variation</td>
</tr>
</tbody>
</table>

*Note.* Black arrows indicate similar ideas. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
Figure 7. Percent of students at each level of construct E. Percent of students at each level of construct E on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q15: Check the box next to the statement you think best explains what a baby bunny would look like. Inter-rater reliability > 85%.
level 3; less than 17% on the pre interviews, Figure 7B, level 3). All other students in context 1 and all students in the contexts 2 and 3 were unable to describe correct simple correlations between genes and traits (Figure 7, levels 0-2). The correlation between genes and traits is the cornerstone of the genetic model. The data indicate that students do not have a firm grasp of this model prior to instruction.

Student achievement on written assessments significantly increased after instruction in contexts 1 and 3 (Figure 2A), but there were no significant increases in student achievement after instruction in either context on the interview responses (Figure 2B). No students were able to achieve an understanding greater than level 3 on either assessment type (Figure 7, levels 4-5). The most complex idea that students were able to describe even after instruction was the idea that alleles randomly assort independent of chromosomes and that each gamete contains only one of the two possible alleles (Figure 7, level 3). Students who held this idea described an understanding more complex than each parent just contributing half of the genetic information (Table 4.6, original level 1) because they described that the alleles in each sex cell are different or that each parent gives one allele. Since many students held that understanding and the understanding is a productive stepping stone, despite not quite aligning with canonical knowledge about how alleles in close proximity on chromosomes travel together, the idea that each sex cell contains one allele which sorts independent of chromosomes was included on the revised construct as a level 3 understanding (Table 4.6).

Level 4 begins to incorporate simple ideas of the meiotic model of genetics (cells go through meiosis and chromosomes are randomly assorted into gametes) and level 5 incorporates more complex ideas of the meiotic model (chromosomes can swap sections
and recombine during meiosis). Even though these ideas were taught in all three contexts, students were unable to integrate the meiotic model with their understandings of the genetic model even after instruction. The results indicate that students see Punnett squares (heredity/genetics) and meiosis (a cellular process) as separate events and do not understand that they are linked. The student data indicate a very strong need for teachers and curriculum to make explicit the connections between heredity/genetics and meiosis so that students can understand the connections between the genetic and meiotic model. It is likely that intermediate ideas exist between the higher levels of this construct (Table 4.6, levels 4-5), however this study did not uncover such ideas and thus cannot suggest data-drive revisions and refinements to the more complex portion of this construct which integrates the genetic and meiotic models.

**Construct F.** This construct focuses on the idea that there are patterns of correlation between genes and traits and that there are certain probabilities with which these occur. This idea was included in the third intervention unit (“Can we engineer a superhuman?”) which was only taught to students in context 1. The patterned correlation between genes and traits and probabilities with which they occur (the genetic model) is normally taught in 10th grade biology; however the intervention unit provided additional inquiry activities to help students achieve the highest level in the construct which included the function of proteins in these relationships (integration of the molecular model). In the original Duncan *et al.* (2009) learning progression, this construct contained three levels: 1-3 (Table 4.7). Upon empirical testing in this study, those levels of the construct were found to be valid and represented ideas that many students had. Lower and intermediate ideas were also found on this construct.
### Empirical Revisions and Refinements of Construct F

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2009)</td>
<td>Todd (2013)</td>
<td>Level Description</td>
</tr>
<tr>
<td>Level 0</td>
<td>Level 0</td>
<td>No understandings of genes or traits</td>
</tr>
<tr>
<td>Level 1</td>
<td>Level 1</td>
<td>We vary in how we grow and function. For a given trait there are variations. Different organisms have different versions of the trait</td>
</tr>
<tr>
<td>Level 2</td>
<td>Level 2</td>
<td>Individuals have two versions for each gene (alleles). Each chromosome in the set carries one version of the gene. There are patterned correlations between the variants of the genes and the resulting trait</td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 3</td>
<td>Organisms get one allele of a gene from each parent, predictable patterns determine the resulting trait variations</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
<th>Todd (2013)</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 3</td>
<td>The gene variants differ in their nucleotide sequence resulting in different or missing proteins that affect our phenotype.</td>
<td>Revised LP</td>
<td>Level 5</td>
</tr>
<tr>
<td></td>
<td>Dominant and recessive genetic relationships can be explained at the molecular level as a consequence of the function and interaction of gene products.</td>
<td>Todd (2013)</td>
<td>Level 5</td>
</tr>
<tr>
<td></td>
<td>Alleles differ in nucleotide sequence affecting proteins which gives trait variation; dominant/recessive relationships can be explained by protein interactions.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prior to instruction, mainly on written assessments, many students were unable to describe any understanding of genes or traits (Figure 8, level 0). This lack of understanding was added to the revised construct as the lower anchor (Table 4.7, level 0). Reviewing the student responses to a specific question in this construct (Q11: check the box next to the statement you think best explains how Bill was able to produce pink-flowered snapdragons from crossing a white-flowered plant with a red-flowered plant) in more depth, over 60% of students in each context were unable to describe on the written assessment any understanding of traits prior to instruction (Figure 8A, level 0). Although a small number of students only in context 2 were able to correctly describe a patterned correlation between genes and traits prior to instruction on the written assessments (7%, Figure 8A, level 3), the percentage of students in each level on the pre interviews in contexts 1 and 2 were very similar, indicating the students in all contexts had similar very basic understandings of the genetic model prior to instruction. This finding is consistent with performance before instruction on construct E which also deals with the genetic model.

After instruction, 30% of students in context 1, 0% of students in context 2, and nearly 8% of students in context 3 were able to correctly describe a patterned correlation between genes and traits on written assessments (Figure 8A, level 3). The increase in student achievement was only significant in context 3 (Figure 2A). During the post interviews, over 50% of students in context 1 and 45% of students in context 3 were able to reach that same level of understanding of the genetic model (Figure 8B, level 3). Although the shift was only significant in context 1 (Figure 2B), the results demonstrate that students have a good understanding of the genetic model after instruction.
Figure 8. Percent of students at each level of construct F. Percent of students at each level of construct F on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q11: Check the box next to the statement you think best explains how Bill was able to produce pink-flowered snapdragons from crossing a white-flowered plant with a red-flowered plant. Inter-rater reliability > 85%.
Some students, however, were not able to correctly explain patterns of inheritance but knew that offspring received traits or genetic material from both parents. These students explained that the traits or genes “mixed” inside of the offspring or that the traits from one of the parents (often the male) “won out” over the traits from the other parent (Figure 8, level 2). This idea was not included in the original progression but is an important stepping stone in understanding how genetic information comes from both parents and results in the traits of the offspring. Thus, this idea was included in the revised construct as a level 2 understanding (Table 4.7).

The highest level of the original progression (original level 3) integrated ideas from both the genetic and molecular models, describing how alleles differ in nucleotide sequence, affecting the proteins to give trait variations and that dominant and recessive relationships can be explained by the interaction of the proteins produced (Table 4.7). Only a very small portion of students were able to integrate the molecular model into their understanding of the construct after instruction. One student in context 1 was able to do so on the written assessment (Figure 8A, level 5) and just over 7% and 4% of students in contexts 1 and 2, respectively, were able to do so on the post interviews (Figure 8B, levels 4-5). Some of these students were able to integrate a small portion of the molecular model with the genetic model, but were unable to fully articulate how interactions at the molecular level explain the correlation between genes (alleles) and traits. Since this idea was an important conceptual shift in being able to integrate the genetic and molecular models, this idea was included in the revised construct as a level 4 understanding (Table 4.7). The data indicate that students understand the genetic model after instruction but have difficulties integrating the molecular model with the genetic
model. These findings are similar to the findings on construct E where students had difficulties integrating the meiotic model with the genetic model.

**construct G.** This construct focuses on the ideas that DNA varies between individuals and species and that changes to the genetic information can change how organisms look and function. For the purpose of assessing student understandings of these ideas, this construct was divided into two sub-constructs: G1 (DNA varies between individuals and species) and G2 (changes to the genetic information can change how organisms look and function). The first idea (G1) was discussed in the last two intervention units (“Why is a Siamese cat colored the way it is?” and “Can we engineer a superhuman?”) while the second idea (G2) was not included in any unit. Both intervention units were taught in their entirety to students in context 1; only selected lessons from unit 2 were taught to students in context 2. These ideas are normally taught in 10th grade biology; however the intervention units provided additional inquiry activities to help students achieve the highest level in the construct.

Each of the new constructs contain ideas from the original singular Duncan *et al.* (2009) construct G which contained three levels: 1-3 (Tables 4.8-4.9). Upon empirical testing in this study, the levels included in the original construct were found to be valid and represented ideas that many students had. Many lower and intermediate ideas that fit in the two new revised constructs were also found. Due to this construct having two ideas and finding that students hold different understandings of each of these ideas, the suggested revisions and refinements in this study include breaking this construct into two constructs. The revisions to each of the new constructs and student achievement in the constructs will now be discussed in detail.
### Table 4.8

**Empirical Revisions and Refinements of Construct G1**

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0</td>
<td>No idea how genotype affects phenotype</td>
</tr>
<tr>
<td>Level 1</td>
<td>Different organisms have different traits/functions</td>
</tr>
<tr>
<td>Level 2</td>
<td>Organisms have different genetic information</td>
</tr>
<tr>
<td>Level 3</td>
<td>Different organisms have different genetic information, even within a species (such as X and Y in boy v. girl humans), DNA variations between individuals can be used for identification</td>
</tr>
<tr>
<td>Level 4</td>
<td>Organisms of the same species have some similar and some different DNA</td>
</tr>
<tr>
<td>Level 5</td>
<td>Some DNA varies between species and some does not (we share some genes with other species such as mice and flies)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level Description</strong></td>
<td><strong>Level Description</strong></td>
</tr>
<tr>
<td>Level 6</td>
<td>Shared DNA codes for things critical to life; the more conserved, the more critical the gene product</td>
</tr>
</tbody>
</table>

*Note.* Strikethrough indicates an idea that was removed from the construct. Black arrows indicate similar ideas. Dashed arrows indicate underlined ideas that were combined into one level. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
Table 4.9

**Empirical Revisions and Refinements of Construct G2**

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan <em>et al.</em> (2009)</td>
<td>Todd (2013)</td>
<td><strong>Level 0</strong> No idea why organisms look different</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Level 1</strong> Different species of organisms look/function differently</td>
</tr>
<tr>
<td>Level 1</td>
<td></td>
<td>Different organisms vary in how they look and function because they have different genetic information. Even within a group of organisms there is variation in traits</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Level 2</strong> Organisms within a species look/function differently because they are different organisms</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td>The genetic information can sometimes change. Changes in the genetic information can result in changes to the structure and function of proteins. Some changes can be beneficial, others harmful, and some neutral to the organism in its environment. Chromosomes, like X and Y, also vary in boys versus girls</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Level 3</strong> Changes to an organism could be beneficial or harmful to an organism</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Level 4</strong> DNA changes could be beneficial/harmful/neutral to an organism; these changes result in changes to the protein structure/function</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Original LP Duncan et al. (2009) Level Description</th>
<th>Revised LP Todd (2013) Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 3 DNA mutations are the source of genetic variation. Some DNA sequences can vary between species while others do not, therefore, we share some genes with other species (mice, flies). DNA sequences can vary between individuals and allow us to differentiate between individuals.</td>
<td>Level 5 DNA changes lead to increased genetic variation and evolution of a species over time (may or may not indicate that population shift occurs towards beneficial trait).</td>
</tr>
</tbody>
</table>

*Note.* Strikethrough indicates an idea that was removed from the construct. Black arrows indicate similar ideas. Italicized text indicates content that was added to a level. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
This construct deals with the idea that DNA varies between individuals and species. In the original Duncan et al. (2009) construct G, all three levels of the construct contained the idea that DNA varies between individuals in a species (Table 4.8, underlined text). Since this was the same idea stated in different ways and upon empirical testing many students were found to hold this idea (Figure 9, level 3), the statements from the different levels were condensed into a single level on the revised construct describing the idea that organisms within a species have different genetic information (Table 4.8, level 3).

Prior to instruction, students in all contexts had moderate understandings about DNA variations between individuals and species in both written assessments (Figure 2A) and in interviews (Figure 2B). Reviewing the student responses to a specific question in this construct (Q10: which of the following statements do you think best explains why the flowering plants look different) in more depth, some students in contexts 1 and 2 were able to achieve the highest level in the progression prior to instruction. On the written assessments, 7% of students in each context were able to explain that some genes are shared between species (Figure 9A, level 5). In the interviews, 13% and 4% of students in contexts 1 and 2, respectively, were able to achieve this level of understanding prior to instruction (Figure 9B, level 5). This indicates that students enter 10th grade with better understandings of this construct than others. Genetic similarities and differences between and within species is often discussed in scientific literature and videos produced for the mainstream public which may explain why students have a better grasp of this construct prior to instruction.
Figure 9. Percent of students at each level of construct G1. Percent of students at each level of construct G1 on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q10: Which of the following statements do you think best explains why the flowering plants look different? Inter-rater reliability > 85%.
Although many students demonstrated an understanding of this concept, especially after instruction (Figure 9, level 5), students also held a variety of other ideas about DNA variation between and within species which prompted further revisions to the construct. Student achievement on written assessments significantly increased after instruction only in contexts 1 and 3 (Figure 2A). Significant increases in student achievement after instruction were also seen in contexts 1 and 2 on the interviews (Figure 2B).

Some students were unable to describe how a genotype relates to a phenotype; this idea was added as the lower anchor of the construct (Table 4.8, level 0). Some students were only able to describe how different organisms have different traits or functions. This idea was included in the original level 1 of the construct (Table 4.8), but the level also contained the additional idea that organisms within a species have different DNA. Since this level contained two ideas and empirical data in this study demonstrated that some students only understand that different organisms have different traits or functions, this idea was separated into a new level in the refined construct (Table 4.8, level 1).

Some students also described the more complex idea that different organisms have different genetic information but did not discuss the idea that organisms within a species may also have different DNA (Figure 9, level 2). Although this may be a small distinction of organisms within or between species having different DNA, when students only describe that different organisms have different DNA, it cannot be assumed they understand that organisms within a single species have variation in their DNA. These students may explain that roses and daisies have different DNA but since plants are a
more unfamiliar context than mammals, the students may think that all roses have the same DNA. Because this idea does represent an important conceptual shift from understanding that traits and functions vary between organisms to understanding that DNA varies between organisms, this idea was included in the revised construct (Table 4.8, level 2). This level also implicitly includes the idea that all living organisms have genetic material, which was an idea originally included in construct A and removed during its revision (Table 4.1). The idea that all organisms have genetic material is better suited for this construct because in order to understand that DNA varies between organisms, students must understand that all organisms have DNA. Students who do not understand that all organisms have DNA will not be able to achieve the understanding that DNA varies between organisms, especially between species.

Particularly during the interviews, many students described that organisms within a species have some similar and some different DNA (Figure 9, level 4). This idea sounds similar to the idea that organisms within a species have different DNA (Table 4.8, level 2); however the ideas are conceptually different. Students who explain organisms within a species have different DNA are only explaining that differences exist in genetic information in organisms of the same species. Students who explain organisms within a species have some similar and some different DNA are explaining that similarities and differences exist in the genetic information in organisms of the same species, thus a more complex and scientifically accurate understanding. The more complex idea (organisms of the same species have some similar and some different DNA) was not originally included in the construct, but since it represents an important conceptual shift in students’ understanding that genetic similarities and differences exist in organisms (albeit, just
within a species), it was included as a new level in the revised construct (Table 4.8, level 4).

Overall, students understood this construct very well. After instruction over 14% of students in context 1, over 7% of students in context 2, and nearly 16% of students in context 3 able to demonstrate understanding of the highest level of the progression on the written assessments (Figure 9A, level 5) and over 60% of students in context 1 and nearly 55% of students in context 2 were able to do in post interviews (Figure 9B, level 5). This data indicate that this construct may be one of the conceptually least difficult for students and suggests that students may be able to achieve an even higher learning performance than originally hypothesized.

Interestingly, while Duncan et al. (2009) did not include a more complex idea in the progression itself, the authors did discuss a more complex idea in their paper describing this construct. They explain that this construct is important for students to learn because the genetic similarities between species is useful for medical and therapeutic research, specifically animal models. The authors point out that the more conserved a particular gene and gene product is, the more important that particular protein is for the function of the organism (Duncan et al., 2009). Given that so many students both before and after instruction were able to explain that organisms between species have similar and different DNA (Figure 9, level 5), it is not unreasonable to think that some students would be able to achieve a higher learning performance if the idea were included in the molecular genetics instruction. Therefore, the revised construct contains an additional higher learning performance, describing how shared DNA codes for entities critical to life and that the more a gene is conserved, the more its gene product
is important (Table 4.8, level 6). Assessment items are needed to test the higher learning performance to determine if students are able to achieve the understanding in different contexts. The large number of students who are able to explain this construct very well suggests that it is not unreasonable to expect students to be able to achieve the learning performance given the idea’s inclusion during instruction.

**G2.** This construct deals with the idea that changes to the genetic information can change how organisms look and function. Prior to instruction, in both written assessments (Figure 2A) and in interviews (Figure 2B), students in all contexts had moderate understandings about how changes to a genotype can change phenotype. Reviewing the student written and interview responses prior to instruction for a specific question in this construct (Q13: which scientist do you think best explained what would happen to the plants if they survived being “fertilized” with the pesticide) in more depth, nearly 62% of students in context 1, 48% of students in context 2, and 70% of students in context 3 were unable to describe how genes relate to traits (Figure 10A, level 0). This idea was included on the revised construct as the lower anchor (Table 4.9, level 0).

Some students also explained the idea that different organisms look and function differently (Figure 10, level 1), but they did not state that organisms within a species can look and function differently. The idea that organisms within a species can look and function differently was included in the original construct and was retained as a level 2 response (Table 4.9). Because some students articulated an idea that was a productive stepping stone to realizing that organisms even within a species can look and function different, the idea that different organisms look and function differently was included in the revised construct (Table 4.9, level 1).
Figure 10. Percent of students at each level of construct G2. Percent of students at each level of construct G2 on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q13: Which scientist do you think best explained what would happen to the plants if they survived being “fertilized” with the pesticide. Inter-rater reliability > 85%.
The students were better able to explain how genes and traits are connected during interviews. A large portion of the students explained that changes to an organism could either be beneficial or harmful; 70% of students in context 1 and 62% of students in context 2 explained this prior to instruction (Figure 10B, level 3). The students were often able to give examples of beneficial and harmful changes. This idea was not included in the original construct but a large number of students did demonstrate this understanding. Additionally, this understanding represents an important conceptual shift where students begin to think about how changes can either be beneficial or harmful to an organism. As such, this idea was included on the revised construct (Table 4.9, level 3).

There was an increase in the number of students who were able to reach levels 4-5 of this construct after instruction, but the only significant gains were seen in the interviews with context 2 students (Figure 2B). A level 4 understanding requires students to understand that DNA changes could be beneficial, harmful, or neutral and that these changes result in changes to the protein structure and function. Only 4% of context 1 students in written assessments were able to reach this level after instruction (Figure 10A, level 4). Nearly 4% of students’ interviews in context 1 and 13% of students in context 2 held this belief after instruction (Figure 10B, level 4).

Genetic changes to organisms drive evolution and natural selection. Evolution and natural selection was not mentioned in the original construct G itself but was mentioned in the text describing the construct (Duncan et al., 2009). Several students were able to explain how beneficial changes to organisms could lead to evolution of a species over time, especially after instruction (Figure 10, level 5). Because the idea that genetic material can be shared between species was moved to construct G1 and the idea
of genetic changes driving evolution was included in the original text describing the construct, the idea was included in the highest level of this revised construct (Table 4.9, level 5). After instruction, 4% of students’ written assessments in contexts 1 and 2 were able to reach this level of understanding (Figure 10A, level 5). No students in context 3 were able to do so. More students were able to explain this idea during interviews. After instruction, 38% of students in context 1 and 22% of students in context 2 were able to explain that DNA changes lead to increased genetic variation and evolution of a species over time. Although those are large numbers, over half of the students interviewed in both context 1 and 2 still held the more basic idea that changes could be beneficial or harmful to organisms (Figure 10B, level 3) even after instruction. This indicates that even after instruction, students have difficulties understanding how genetic changes drive evolution and natural selection.

**construct H.** This construct focuses on the idea that environmental factors can interact with genetic information. This idea was included in the first intervention unit (“How do cells become cancerous?”) which was taught to students in contexts 1 and 2. Environmental influence on genetics is typically taught in 10th grade biology at a very basic level, however the intervention unit provided additional inquiry activities to help students understand that the environment can mutate genes which alter proteins or their expression. In the original Duncan *et al.* (2009) learning progression, this construct contained three levels: 1-3 (Table 4.10). Upon empirical testing in this study, those levels of the construct were found to be valid and represented ideas that many students articulated. Some lower and intermediate student ideas were also documented on this construct.
### Table 4.10

**Empirical Revisions and Refinements of Construct H**

<table>
<thead>
<tr>
<th>Original LP</th>
<th>Revised LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 0</strong></td>
<td>Level 0</td>
<td>No understanding of environmental impact on organisms</td>
</tr>
<tr>
<td><strong>Level 1</strong></td>
<td>Level 1</td>
<td>The environment cannot affect organisms (at any of the trait/cell/DNA levels)</td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td>Level 2</td>
<td>The environment can affect our traits or functions</td>
</tr>
<tr>
<td>Level 1</td>
<td>The environment can affect our traits. Even organisms that are related may end up looking or behaving differently</td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td>The environment can influence cell function through changes at the protein level (type and amount)</td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 3</td>
<td>The environment can affect our cells or organs or tissues</td>
</tr>
<tr>
<td>Level 4</td>
<td>Level 4</td>
<td>The environment can change (or mutate) things (such as DNA, genes, molecules, etc.) inside of the cell</td>
</tr>
<tr>
<td>Level 5</td>
<td>Level 5</td>
<td>The environment can change proteins (type and amount) which influence cell function</td>
</tr>
</tbody>
</table>

(continued)
| Original LP |  | Revised LP |  |
|------------|  |------------|  |
| Duncan et al. (2009) | Level Description | Todd (2013) | Level Description |
| Level 3 | Environmental factors can cause mutations in genes, or alter gene expression | Level 6 | The environment can mutate genes which change proteins or alter gene expression of proteins |

*Note.* Black arrows indicate similar ideas. Grey background indicates an idea that was added to the construct. Parts adapted from “A Learning Progression for Deepening Students’ Understandings of Modern Genetics Across the 5th - 10th Grades,” by R. G. Duncan, A. D. Rogat, and A. Yarden, 2009, *Journal of Research in Science Teaching*, 46(6), pp. 660-661. Copyright 2009 by Wiley Periodicals, Inc.
Students’ written pre assessments and pre interviews in all contexts revealed simple understandings of the impact of the environment on genes (Figures 2A, 2B). Prior to instruction, mainly on the written assessments, some students were unable to reason if the environment had any impact on organisms (Figure 11, level 0). Because this idea was not included in the original Duncan et al. (2009) progression and several students expressed this idea, it was included in this construct as the lower anchor (Table 4.10, level 0). Prior to instruction, a small number of students explained that the environment cannot affect organisms (Figure 11, level 1). Even though this idea is incorrect, it was included in the revised construct (Table 4.10, level 1) because the idea directly relates to the content and demonstrates an important conceptual shift in student understandings of the environmental impact on organisms. Thinking that the environment cannot impact organisms is a very different conceptual idea from the more complex idea that the environment can affect an organism’s traits or functions (Table 4.10, level 2). Knowing which students do not think environment can affect traits is useful instructional leverage because teachers could provide specific examples of how the environment can change organisms at different levels of sophistication to specific students.

Reviewing the student responses to a specific question in this construct (Q12: check the box next to the statement you think best explains what would happen to the plants after they were exposed to the pesticide) in more depth, only 9% of students in context 1, 10% of students in context 2, and 13% of students in context 3 were able to explain in writing before instruction that the environment can influence things at the cellular or subcellular level (Figure 11A, levels 3-6). More students were able to explain this idea before instruction in the interviews; 30% of students in context 1 and 29% of
Figure 11. Percent of students at each level of construct H. Percent of students at each level of construct H on pre and post written assessments (A) or pre, middle, and post interviews (B). Responses on both written assessments and interviews are from Q12: Check the box next to the statement you think best explains what would happen to the plants after they were exposed to the pesticide. Inter-rater reliability > 85%.
students in context 2 understood that the environment can influence entities at the cellular or subcellular level (Figure 11B, levels 3-6).

Significant increases in student achievement were only seen in context 1 on both written assessments (Figure 2A) and interviews (Figure 2B). No significant increases in student understandings were seen in either context 2 or 3. After instruction, 75% of the students in context 1 and 50% of the students in context 2 believed that the environment could influence entities at the subcellular level (Figure 11B, levels 4-6). Although the number of students increased in each context, the only significant change was in context 1. Students in context 1 were also much more able to explain that the environment can change genes which can alter proteins or their expression, which is the highest level on this construct. After instruction, 10% of students on the written assessments and 34% of students in the interviews in context 1 were able to explain the idea while no students on the written assessment and 4% of students on the interviews in context 2 were able to explain the idea (Figure 11, level 6). This indicates that students can reach this level in the construct, but that it is difficult for students to understand that the environment can change proteins or alter their expression.

Two levels were added to this construct between the original levels 1 and 2 (now revised levels 2 and 5). The revised level 3 describes the idea that the environment can affect organisms at the cell, organ, or tissue level while the revised level 4 describes the idea that the environment can change entities in an organism at the subcellular level (Table 4.10, levels 3-4). Students held these two ideas both before and after instruction on both written assessments and interviews (Figure 11, levels 3-4). Adding levels corresponding to changes and the cellular level and changes at the subcellular level are
also consistent with the revisions to constructs B and C (Tables 4.2-4.4) because they describe important conceptual shifts in student understanding of changes at different organizational levels inside organisms.

**combining the Duncan et al. (2009) and Roseman et al. (2006) progressions.**

The Duncan et al. (2009) learning progression was chosen for data analysis over the Roseman et al. (2006) progression because the Duncan progression is divided into eight “Big Ideas” and also describes a learning performance for three different levels for each of the “Big Ideas,” making mapping student achievement in the different ideas easier. The Duncan et al. (2009) progression will be used as a template for combining the two molecular genetics learning progressions into one. Although both progressions are for molecular genetics content, there are differences between the progressions in terms of content included. Since the content included in the Duncan et al. (2009) progression has been discussed in detail in the previous sections, only similarities and differences in content between the progressions will be presented in addition to how information from the Roseman et al. (2006) progression can be used to further modify the revised Duncan et al. (2009) progression.

Of the 23 ideas included in the Roseman et al. (2006) progression, all but five of the ideas can be mapped to learning performances in the Duncan et al. (2009) progression (Figure 12). Each “Big Idea” construct from the Duncan et al. (2009) progression is represented by at least one idea in the Roseman et al. (2006) progression; that is to say, each construct in the Duncan et al. (2009) progression is also represented in the Roseman et al. (2006) progression, although the constructs are not separated and learning performances do not have defined levels.
Three of the ideas in the Roseman et al. (2006) progression involve fundamental ideas of biology and chemistry that are integral to understanding molecular genetics, but may not be necessary to include in a progression for molecular genetics (Figure 16, lower three boxed ideas). The idea that “all matter is made up of atoms... atoms may stick together in well-defined molecules or may be packed together in large arrays. Different arrangements of atoms into groups compose substances.” (Roseman et al. 2006, p. 6), is very important for understanding how changes to DNA bases can change proteins, how the environment can damage DNA, and how proteins and enzymes do work in the cell, among other ideas. However, this idea is more background knowledge of basic biology and chemistry concepts. Although important for understanding molecular genetics, this idea should be included on learning progressions for biology and/or chemistry because it is not directly related to specific molecular genetics content.

Similarly, the ideas that “Some living things consist of a single cell. Like familiar organisms, they need food, water, and air; a way to dispose of waste; and an environment they can live in.” and “Within cells, many of the basic functions of organisms–such as extracting energy from food and getting rid of waste–are carried out. The way in which cells function is similar in all living organisms.” (Roseman et al., 2006, p. 6) are both fundamental ideas of biology regarding the function of cells and important for understanding similarities and differences between organisms, but the ideas are not directly related to content in molecular genetics. The idea that organisms between species have similarities is included in construct G1 (Table 4.8), but it does not discuss the need cells have for food, water, and air; a way to dispose of waste; and an environment in which they can live because these ideas are not related specifically to
molecular genetics. They are general concepts in biology as a whole, but not molecular genetics. The idea that proteins provide functions inside cells is included in construct C1 (Table 4.3), but like previously discussed, the necessities for cells to live are not discussed because the necessities are not related directly to molecular genetics. Even though these three ideas are important for understanding concepts in molecular genetics, they do not deal directly with concepts specifically in molecular genetics. Therefore, none of these three ideas were included in the modifications to the Duncan et al. (2009) progression.

The ideas that “An altered gene may be passed on to every cell that develops from it (that cell)” and “when mutations occur in sex cells, they can be passed on to all cells in the resulting offspring, if mutations occur in other cells, they can be passed on to descendent cells only” (Figure 12, upper two boxed ideas) were also not directly included in the Duncan et al. (2009) learning progression. These ideas involve the meiotic model of genetics, or how genetic information passes from generation to generation. Construct E (Table 4.6) discusses the basic ideas of meiosis (each parent contributing half of the genetic information through independent assortment of chromosomes and genetic variation through recombination), but does not include the impact of genetic changes. Construct C2 (Table 4.4) discusses the impact of genetic changes to protein structure and function but does not include how changes may or may not be passed on to offspring. Construct G2 (Table 4.9) discusses how DNA changes can result in genetic variation and evolution of a species over time but does not include that only mutations in gametes can be passed to offspring.
Because these ideas are directly related to molecular genetics content and were included in the Roseman et al. (2006) progression, these ideas could be added to the Duncan et al. (2009) progression as a new construct to begin to combine the two progressions. A hypothetical construct, I, containing these ideas is shown in Table 5. Although the levels in this construct are based on student understandings found in the literature (Bowling, et al., 2008; Smith, Wood, & Knight, 2008; Tsui & Treagust, 2010), the Roseman et al. (2006) progression, and during conversations with students involved in this study, this construct remains a hypothetical model of students learning since this construct has not been empirically tested in any classrooms.

A level 0 describes students who are unable to explain how mutations could be passed on to offspring and is the lower anchor of the construct. Level 1 describes understandings that traits can be passed on to offspring, such Lamarck’s incorrect theory of inheritance of acquired characteristics (that giraffes developed long necks over time because they stretched to reach leaves at the tops of trees or that a man who lifts weights and is very muscular can pass along large muscles to his children). Although these are incorrect ideas, they represent an important stepping stone to understanding correctly how changes can be passed on to offspring. A level 2 understanding describes a more complex idea that only changes to DNA can be passed on to offspring. Students who explain this idea understand that physical traits themselves cannot be passed down, only DNA. Thus, only changes to the DNA can be passed on to offspring. Level 3 describes the understanding that only DNA changes to gametes can be passed on to offspring. Students at this level understand that only the genetic information contained in gametes is passed on to offspring, thus any changes to somatic cells (like the development of skin
Table 5

_Hypothetical Construct 1_

<table>
<thead>
<tr>
<th>Hypothesized LP</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0</td>
<td>Unable to describe how mutations could be passed on to offspring</td>
</tr>
<tr>
<td>Level 1</td>
<td>A change in traits of an organism can be passed on to offspring</td>
</tr>
<tr>
<td>Level 2</td>
<td>DNA mutations in organisms can be passed on to offspring</td>
</tr>
<tr>
<td>Level 3</td>
<td>Only DNA mutations in gametes can be passed on to offspring</td>
</tr>
<tr>
<td>Level 4</td>
<td>DNA mutations that occur in gametes can be passed on to offspring resulting from that gamete while DNA mutations that occur in somatic cells can be passed on to descendend cells only</td>
</tr>
</tbody>
</table>
cancer later in life due to an accumulation of mutations) is unable to be passed on to the resulting offspring. A level 4 understanding is the upper anchor of the construct and describes how changes to gametes can be passed on to offspring resulting from the changed gamete and that changes to somatic cells can be passed on to descendent cells as a result of mitosis.

The Roseman et al. (2006) and Duncan et al. (2009) molecular genetics learning progressions contained many of the same content ideas, however some ideas were mutually exclusive. By including the ideas exclusive in the Roseman et al. (2006) progression that were directly related to molecular genetics in the revised Duncan et al. (2009) progression as a new construct, a more complete progression is created. The more complete single progression is useful for researchers as well as teachers when trying to design curriculum and instruction to targeted aspects of molecular genetics in order to increase student understandings of the ideas in molecular genetics.

**Impact of Intervention Units**

Since neither the Roseman et al. (2006) nor the Duncan et al. (2009) molecular genetics learning progression has yet been fully validated and neither contains any instructional or curricular supports, this study aimed to determine the impact of three intervention units to determine if intervention units can promote student progress through a learning progression. If these units and assessments are found to help support student progress, they could then be added to the progressions. As instructional and curricular supports are added to the progressions, the more practical and useful the progressions become for teachers and researchers. It was hypothesized that activities in the intervention units that target specific constructs in the Duncan et al. (2009) progression
will help students achieve higher learning performances in those constructs compared to students who are not exposed to the intervention units. This simplistic finding was not the case, likely due to school and classroom culture, and teacher effects.

Students in classroom contexts 1 and 2 received intervention units while classroom context 3 did not. Context 1 completed all three of the units in their entirety and context 2 completed all of intervention unit 1 and a shortened version of intervention unit 2. The three units targeted constructs B, C, D, F, G, and H of the Duncan et al. (2009) learning progression. As shown in Figure 2, students in context 1 showed a greater understanding of each construct after instruction than students in either of the two other contexts, indicating that the three intervention units may have helped increase student achievement. However, students in context 2, who received one and a half intervention units, showed the lowest achievement on written assessments across nearly all constructs, with significant gains only in construct A which was not targeted by an intervention unit (Figure 2A). During interviews, students in context 3 did show significant gains in some of the constructs (Figure 2B), including constructs B, C, D, and G targeted with the intervention units, but not to the same level as students in context 1 who also received intervention units. These findings underscore that student achievement is based on a combination of factors including school and classroom culture, teacher effects (Draper 2010), curriculum and instruction, poverty, parental involvement, and many other factors beyond control of simple presence or absence of intervention units.

To try to determine what impact, if any, the intervention units did have on student achievement, individual student responses from students in context 1 and 2 were analyzed. Responses to both written and interview questions that included an explicit
reference to an idea mentioned in an intervention unit were noted. Responses from students in context 3 were not analyzed since they were not exposed to the intervention units, thus never had an opportunity to reference ideas in the units. In order to be noted, the reference had to refer to information that came directly from the intervention units. General references to cancer were not counted because the teacher in context 1 did other units on cancer and the students could have been referring to those units in other teacher-produced units besides the molecular genetics intervention unit 1.

Students in context 1 made explicit references to ideas presented in the intervention units when responding to questions in constructs B, C, D, F, G1, G2, and H, the constructs the intervention units targeted. Students in context 2 made explicit references to ideas presented in the intervention units when responding to questions in constructs B, C, D, G2, and H, five of the seven constructs the intervention units targeted. Since construct F was only targeted in intervention unit 3, it was not surprising that students in context 2 did not make any references to intervention unit 3 ideas as they did not complete the intervention unit. After noting the responses that contained an explicit reference to an idea or ideas presented in the intervention units, the responses were categorized according to the type of reference. The two main categories of references the students discussed were examples of protein structure and function (discussed in intervention units 1-2), and the “gene, protein, cell, trait” (GPCT) scaffold originally described by Duncan et al. (2011) included in intervention unit 3. Since the GPCT scaffold was only included in intervention unit 3, only students in context 1 made references to this idea while answering questions.
For each context, references were analyzed according to students’ explicit references versus students’ lack of references. Only four students in context 2 made any explicit references to protein structures and functions included in the first two intervention units, such as the role of hemoglobin in red blood cells or how proteins can be denatured and not perform their functions. The four students referenced an example of protein structure and function included in interviews during probing of the four completely different contexts: B, C, G, H. Each of the four responses that included a reference to a protein structure and function included in the intervention units was at least a level 3 response, indicating at least moderate understanding of the construct. One response cannot be compared to all other responses of the construct to determine impact of the intervention units in this context. Because of the few number of context 2 students who made explicit references to ideas included in the intervention units, the impact of the units in classroom context 2 cannot be determined.

More students in context 1 referenced ideas mentioned in the intervention units on both written assessments and in interviews, so responses from students in context 1 who mentioned the ideas could be compared to responses from students in context 1 who did not mention the ideas. The students who referenced protein structure and function ideas from the intervention units in their responses to a construct or constructs were compared to students who did not make any references to protein structure function included in the intervention units to determine any significant differences in achievement based on including specific examples of protein structures and functions. Another group of context 1 students who referenced the GPCT scaffold in their responses to a construct or constructs were also compared to context 1 students who did not reference the GPCT
scaffold to determine any significant differences in achievement based on using the GPCT scaffold.

**protein structure and function.** Nine interview students in context 1 made two or more explicit references to protein structure and function ideas mentioned in the intervention units. When the nine students’ responses were compared to the other 25 interview students’ responses in each construct, there was only one statistically significant difference in student achievement between the groups. Construct B explains the idea that genes code for protein structure and function (Table 3.2) and the 9 students who made 2 or more explicit references to protein structures and functions mentioned in the intervention units did significantly better at explaining this construct during the middle interviews than the other 25 students (Figure 13). Both groups of students have very similar levels of understanding of this construct prior to instruction (Figure 13, pre interview) and after all molecular genetics instruction (Figure 13, post interview). The difference in achievement occurred during the middle interview, indicating that the students who were able to make two or more references to protein structures and functions mentioned in the intervention units were more quickly able to grasp the idea that genes code for protein structure and function. Both groups of students demonstrated increased understanding of this construct during the middle interview compared to their pre interview (Figure 13); however the 9 students who made two or more references to protein structure and function ideas mentioned in the intervention units had a significantly higher increase in understanding than students who made less than two references. Both groups of students ultimately have a high level of understanding of this construct after instruction (Figure 13, post interview), but the increased understanding demonstrated in
Figure 13. Average of interview responses for construct B. Average of student responses for the pre, middle, and post interview question for construct B for students who made 0-1 explicit references (squares) or 2 or more explicit references (circles) to protein structure and function ideas included in the intervention units. Responses are to interview Q3: Check the box next to the statement you think best explains how DNA is involved in muscle function. Error bars represent standard deviation. Two sample unequal $t$-test, $^* p < 0.0055$. Inter-rater reliability > 85%.
the middle interview may indicate that providing students with concrete examples of structures and functions of proteins helps students understand the idea that genes code for protein structure and function.

**gene, protein, cell, trait scaffold.** Five interview students in context 1 made one or more explicit references to the gene, protein, cell, trait (GPCT) scaffold mentioned in intervention unit 3; of the five students, three made two or more references to the GPCT scaffold. When these five students’ responses were compared against the other 29 interview students’ responses in each construct, there were three statistically significant differences in student achievement between these groups in constructs D, G1, and G2.

Construct D describes the idea that all cells have the same DNA but expression of proteins is specific for certain cell types (Table 3.4). The two groups of students (those who made no explicit references to the GPCT scaffold and those that did) began the year with very similar understanding of this concept (Figure 14, pre interview). As time went on, the two groups began to diverge with those who explicitly mentioned the GPCT scaffold articulating a better understanding of this construct at both the mid and post interviews, although only significantly better at the post interview (Figure 14, mid and post interviews). The difference in achievement may indicate that a scaffold that has students explain the role of proteins in connecting genes and traits is helpful for students understanding that the expression of proteins is specific for certain types of cells. However, it is interesting to note that not one student made reference to the GPCT scaffold while answering any questions related to construct D although the GPCT scaffold is relevant to this construct, particularly the GPC (gene, protein, cell) portion. Because no students referenced the GPCT scaffold in any answers for construct D, it
Figure 14. Average of interview responses for construct D. Average of student responses for the pre, middle, and post interview question for construct D for students who made 0 explicit references (squares) or 1 or more explicit references (circles) to the gene, protein, cell, trait scaffold (GPCT) included in the third intervention unit. Responses are to interview Q7: Which student do you think best explained why the cells [skin and nerve] looked different. Error bars represent standard deviation. Two sample unequal \( t \)-test, * \( p < 0.0055 \). Inter-rater reliability > 85%.
appears that the students who made the scaffold references understood the construct better for reasons other than the presence of the scaffold.

Construct G1 describes the idea that genes vary between individuals and species (Table 3.7). Prior to instruction, the students who made one or more explicit references to the GPCT scaffold in mid and post interviews articulated a better understanding of this construct than the students who did not make any references to the GPCT scaffold in subsequent interviews (Figure 15, pre interview). Although this difference in achievement on the pre interview was not statistically significant, there was a difference in understanding prior to instruction. In the subsequent interviews, students who made one or more references to the GPCT scaffold did have a statistically significantly better understanding of this construct during and after instruction (Figure 15, mid and post interviews). Because of the difference in understanding prior to instruction, even though it was not a significant difference, it is difficult to say if the GPCT scaffold helped the students have a better understanding of how genes vary between individuals and species or if identification of students who referenced the GPCT scaffold happened to also select students who were inclined to perform better on this construct regardless of the GPCT scaffold.

Construct G2 describes the idea that genetic changes drive evolution and natural selection (Table 3.8). There was no statistically significant difference in student achievement between students who made one or more references to the GPCT scaffold and students who made no references to the scaffold. However, there was a difference between students who made two or more references to the scaffold and students who made 0-1 references. There was a very slight difference in understandings of this
Figure 15. Average of interview responses for construct G1. Average of student responses for the pre, middle, and post interview question for construct G1 for students who made 0 explicit references (squares) or 1 or more explicit references (circles) to the gene, protein, cell, trait scaffold (GPCT) included in the third intervention unit.

Responses are to interview Q10: Which of the following statements do you think best explains why the flowering plants look different? Error bars represent standard deviation. Two sample unequal t-test, * $p < 0.0055$. Inter-rater reliability > 85%.
construct prior to instruction between the two groups of students (Figure 16, pre interview), but this slight difference was not statistically significant. During and after instruction, students who made 2 or more references to the GPCT scaffold had a better understanding of this construct than students who made 0-1 references to the scaffold, although only significantly better on the post interview (Figure 16, mid and post interviews). Because of the small number of students in one of the groups (three students in the 2 or more GPCT scaffold references group) and absence of the molecular model from this construct, it is likely that the small group of students who made two or more references to the scaffold were just more inclined to understand this construct better, regardless of the GPCT scaffold which explains the role of proteins connecting genes to traits.

Overall, it appeared that ideas from the intervention units were most useful for students in context 1, although some students from both contexts did make explicit references to ideas from the intervention units. The examples of protein structure and function included in intervention units 1 and 2 were most useful for students to explain the idea that genes code for proteins (construct B). The GPCT scaffold included in intervention unit 3 originally described as an intervention developed for middle school students by Duncan et al. (2011) seemed the most useful to students as they used it to help answer questions in constructs C, F, G1, G2, and H. Significant differences in student achievement were seen between groups of students who made at least one reference or no references to the GPCT scaffold in constructs D and G1 (Figures 14-15) and between groups of students who made at least two references or 0-1 references to the GPCT scaffold in construct G2 (Figure 16). However, it is unclear if these differences
Figure 16. Average of interview responses for construct G2. Average of student responses for the pre, middle, and post interview question for construct G2 for students who made 0-1 explicit references (squares) or 2 or more explicit references (circles) to the gene, protein, cell, trait scaffold (GPCT) included in the third intervention unit. Responses are to interview Q13: Which scientist do you think best explained what would happen to the plants if they survived being “fertilized” with the pesticide. Error bars represent standard deviation. Two sample unequal t-test, * $p < 0.0055$. Inter-rater reliability > 85%.
are actually due to the GPCT scaffold itself or if the groups of students who referenced the scaffold were students who were likely to perform better after instruction, regardless of scaffold inclusion.
IV. Discussion

Empirically Testing and Revising the Molecular Genetics Learning Progressions

Student understandings of molecular genetics were probed in three different 10th grade contexts. It was expected that students would fall on the extremely low levels of achievement and learning performances before classroom instruction and progress to the higher levels of achievement as the instructional period progressed, allowing student progress to be tracked through the entire progression. It was also expected that students would hold several ideas not included in the original progressions and that these ideas could be used to revise and refine the progressions. Student understandings did align with the Duncan et al. (2009) progression, the students did fall on the extremely low levels of achievement and learning performances before classroom instruction, and students did hold intermediate ideas that were used to modify the progressions.

The main testing and revisions to the molecular genetics learning progressions were done using the Duncan et al. (2009) progression because it included eight defined constructs with three learning performances described for each construct. Constructs A, D, E, F, G, and H (Tables 4.1, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10) had not yet been revised based on empirical data from any studies. Constructs B and C (Tables 4.2, 4.3, 4.4) were recently revised based on empirical data obtained from one middle school classroom (Shea & Duncan, 2013).

Construct A deals with the idea that genetic information is hierarchically organized. Students in all three contexts made significant gains after instruction and had
a much better understanding of how genetic information is organized inside of cells. The proposed revisions to this construct added a lower anchor corresponding to no understanding of how genetic information is organized, added intermediate levels that were important conceptual stepping stones for students, and moved the idea that various organisms have DNA as their genetic material to construct G1 (Table 4.1).

Constructs B, C, and D deal with the molecular model of genetics, more specifically that DNA codes for proteins (B), proteins have important functions inside of cells and are the mechanism that connects genes and traits (C), and expression of proteins is specific to a specialized cell’s function (D). Students entered the year with a very minimal understanding of proteins, which is consistent with other studies reported in the literature. Students in all three contexts did make significant gains after instruction, showing that 10th graders are able to understand and explain the molecular model of genetics after instruction. Construct D had not yet been revised and proposed revisions included adding four additional levels based on student data that represented productive stepping stones (Table 4.5).

Constructs B and C were recently revised by the Duncan lab (Shea & Duncan, 2013). Proposed revisions to construct B included combining two levels corresponding to the ideas that genes are non-informational in nature into a single level (Table 4.2). Very few students held this idea before or after instruction (Figure 4, level 1); it is extremely unlikely that data was lost by combining the passive versus active non-informational distinction. Duncan herself even indicated that she was not sure how useful it is to tease students with active versus passive view apart because both understandings are non-informational in nature (Ravit Golan Duncan, personal 153
communication). Overall, the revisions to construct B were minimal and closely aligned with the changes described by Shea & Duncan (2013).

Although construct C was also recently revised by Shea & Duncan (2013), proposed revisions were much more extensive to this construct. Shea & Duncan’s (2013) revisions to construct C refocused the construct around the idea that proteins are central to the functioning of organisms (Table 4.3), which was a part of the original construct C. When assessments were created for this study, the assessments for construct C focused on the idea that proteins are the mechanism that connects genes to traits, the other central idea included in the original construct C. Since there seemed to be two ideas in the original construct C, revisions to this construct include breaking the construct into two constructs: C1 (proteins have a central role in the functioning of organisms, Table 4.3) and C2 (proteins are the mechanism that connect genes and traits, Table 4.4).

While the idea that proteins have a central role in the functioning of organisms was not specifically addressed in this study, when talking with students about proteins and hearing them discuss protein functions, the revised levels of C1 proposed by Shea & Duncan (2013) are ideas that students have and likely represent valid levels of understandings students have in the three contexts in this study (Table 4.3). Additional assessment items addressing this idea would need to be created to further validate Shea & Duncan’s (2013) findings, but anecdotal findings during this study support their revisions.

Along with creating a new C2 construct dealing entirely with proteins being the mechanism that connects genes to traits, several levels of intermediate understandings were added to the construct. The new levels represented ideas from empirical testing that directly related to the construct, were important conceptual shifts in student
understanding, and were productive stepping stones for students to be able to achieve the expert level understanding of the construct.

Construct E explains the idea that organisms transfer their genetic information to offspring. It includes ideas from the genetic and meiotic models. After instruction, students were able to describe the genetic model, explaining simple dominant/recessive relationships with Punnett squares, but students were unable to integrate the meiotic model into their explanations. Proposed revisions to this construct were the addition of three new levels, each corresponding to a productive stepping stone.

Construct F also includes two models (genetic and molecular) and explains the idea that there are patterns of correlation between genes and traits and that these patterns can be explained by interactions between proteins. Like construct E, students were able to explain the genetic model after instruction, this time by explaining codominance, but they were unable to integrate the molecular model into their explanations. Proposed revisions to this construct also included the addition of three new levels corresponding to productive stepping stones.

The findings in constructs B-D, E, and F suggest that students can understand and explain the three models of genetics (genetic, meiotic, and molecular) but have great difficulties integrating the models. Stewart et al. (2005) contended that literacy in molecular genetics consists of understanding the three interrelated models and also knowing how to integrate the models. This research shows that after instruction in biology, 10th graders increase their understandings of the three models but fail to integrate the models for more complex understandings of the field of molecular genetics.
Construct G explains the idea that species share similar DNA and that changes can change phenotype which drives evolution and natural selection. This construct had not yet been revised based on empirical data from any study. Based on data obtained in this study, revisions to this construct included dividing the construct into two constructs, G1 (DNA varies between individuals and species) and G2 (changes to the genetic information can change how organisms look and function), adding levels to each of the new constructs, and making extensive re-arrangements of concepts (Tables 4.8-4.9). The construct was divided early in the study, so student understandings of each of these separate ideas were assessed in all three of the classroom contexts. Students were very successful in understanding genetic similarities and differences between and within species after instruction (G1, Figure 9) but had difficulties understanding and explaining how changes to the genetic information drive evolution (G2, Figure 10).

Construct H deals with the idea that the environment can change genetic information, and thus, proteins and expression of proteins. Students were able to explain that the environment can influence entities at the cellular and subcellular level after instruction, but few students were able to explain that the environment can alter protein type and amounts through changes to the DNA. Proposed revisions to this construct included adding four new levels corresponding to productive conceptual stepping stones.

The Roseman et al. (2006) progression contained five ideas that were not included in the original Duncan et al. (2009) progression. Three of the ideas were very basic chemistry and biology ideas. While necessary to understand concepts in molecular genetics, the chemistry and biology ideas were not specific to molecular genetics content, so they were not included in revisions to the progressions. The other two ideas were
directly related to content in molecular genetics and thus were included in a proposed new construct, construct I (Table 5), centered around the idea that genetic mutations to somatic cells can only be passed on to descendent cells in the body while genetic mutations to gametes can be passed on to offspring. Although the levels in this construct are based on student understandings found in literature (Bowling, et al., 2008; Smith, et al., 2008; Tsui & Treagust, 2010), the Roseman et al. (2006) progression, and during conversations with students involved in this study, this construct remains a hypothetical model of students’ learning since this construct has not been empirically tested in any classrooms. The addition of this new construct aligns the content of the two molecular genetics learning progressions into a singular progression.

Overall, students increased performance in each construct after instruction and were able to explain the three models of genetics; however students had difficulties integrating the three models to reach more complex understandings of molecular genetics. This indicates a need for future curriculum and instruction to focus on the integration of the three models. Future research studies could assess integration of the three models and provide teachers with curriculum or other instructional materials that require students to integrate the three models, making the connections more apparent for students.

Although extensive revisions and refinements of the molecular genetics progressions were done during this study, further revisions and refinements still need to be completed. The new levels added to the progression need to be tested in other classroom contexts to determine if they are valid in other contexts, particularly in constructs C1, C2, G1, and G2, which were constructs that were divided. Hypothetical
construct I needs to be tested in any classroom because it remains hypothetical until it is empirically tested in classrooms to determine if students do hold the understandings hypothesized in the construct. Additionally, it may be necessary to add additional constructs related to current content research in molecular genetics. For example, the field of epigenetics is absent in both progressions but scientists are still trying to understand how modifications to histones and DNA help control gene expression and phenotypes of organisms. Since epigenetics is not very well understood in the scientific community, it remains unclear how much, if any, epigenetics content should be included in a learning progression targeted to grades 5-10, though Duncan herself acknowledged the omission of this field in her progression (Ravit Golan Duncan, personal communication).

Future research should also include defining contingencies between the constructs of the molecular genetics learning progression. Defining contingencies between all the constructs was beyond the scope of this research project, but understanding the relationships between the constructs is an important step towards validating the progression. The Roseman et al. (2006) progression may be very helpful for future research studies because it displays ideas now mapped to the Duncan et al. (2009) progression connected by arrows in the style of Project 2061’s Atlas of Science Literacy (AAAS, 2001). Identifying the contingencies between and interrelatedness of the constructs is no trivial task, but it is important for learning progression validation.

Complete learning progressions contain instructional materials and assessments which target specific constructs of the progressions and have been shown to help and assess student achievement of the learning performances. Future research could build
upon the assessments described in this study as well as the intervention units targeting some of the constructs. Assessment items (both written and interview questions, Appendix A, B) probe nine of the now eleven constructs and do so for nearly all of the levels included in each of the revised constructs (construct G1, level 6 being a notable exception, Table 4.8). The assessment items could be modified by future studies to include all levels and constructs of the revised progression. Additionally, specific ideas included in the intervention units could be used by future researchers to craft instructional materials targeting the highest learning performances of each of the constructs.

**Impact of Intervention Units**

Although not a comparison study, it was hypothesized that students in contexts that received intervention units targeting specific constructs of the Duncan *et al.* (2009) progression would achieve higher learning performances in the targeted constructs than students in the context that did not receive the units. However this simplistic finding was not the case. Context 1 completed all three of the intervention units targeting the Duncan *et al.* (2009) learning progression constructs B, C, D, F, G, and H; context 2 completed the first and part of the second intervention units targeting constructs C, D, H, and G; context 3 did not complete any intervention units. Students in context 1 demonstrated the largest gains of the three different contexts, both on the written assessments (Figure 2A) and the interviews (Figure 2B). Since the context 1 teacher taught the three intervention units in their entirety, it was expected that these students would perform better on constructs that were addressed specifically in the intervention units. The students in context 1 did perform better than the other students on the targeted constructs and the other constructs that were not specifically addressed in the intervention units.
Since the teacher in context 2 taught the first intervention unit and a few lessons of intervention unit 2, it was expected that the students in context 2 would perform better than the students in context 3 on the constructs that were addressed in the first two intervention units (C, D, G1, H). Although the context 2 students demonstrated significant gains in C, D, and G1 in interviews (Figure 2B), there were no significant gains in any of these constructs on their written assessments. Additionally, their learning performances on each of these constructs were lower than that of students in context 3 on the written assessments (Figure 2A).

It is impossible to make direct comparisons between the different classrooms due to teacher effects; classroom, school, and district contexts; different students; and different curriculum. However, the finding that students in context 3 performed better than students in context 2 on the written assessment items pertaining to the constructs addressed in the intervention units is very interesting. First, and most importantly, it confirms that students are able to significantly increase their knowledge of molecular genetics with normal classroom instruction. The students in context 3 significantly increased their learning performances after normal classroom instruction in molecular genetics in 6 of the 9 constructs: A, B, D, E, F, G1 (Figure 2A). The six constructs encompass all three models of molecular genetics, so the significant increase indicates that students are able to make progress in all three models with typical classroom instruction.

As reviewed in Draper (2010), a large issue in educational research in general is teacher effects. He explained that the effects of different teachers are nearly always bigger than the effects of different treatments. Teacher and school effects were likely
seen in this study as well. Even though the teacher in context 3 did not implement the intervention units, she was able to instruct her students about the constructs and the students were able to significantly increase their learning performances in 6 of the 9 constructs (Figure 2A). Conversely, even with the provided intervention units, students in context 2 were unable to make significant increases on written assessments after instruction in 8 of the 9 constructs. They were able to make significant increases in 6 of the 9 constructs during interviews, however (Figure 2B). The students who received all three of the intervention units in their entirety (context 1) out performed the students in the other two contexts on both the written assessments and interviews (Figure 2). This is likely due to a combination of teacher effects, school effects, and the intervention units.

During classroom observations, teachers in contexts 1 and 2 both reviewed content, tried to engage students, and tried to make connections between content ideas. However, it was noted that the teacher in context 1 made more connections between ideas and more often indicated how prior content related to current content being taught. The teacher in context 2 gave an impression to the students that the content was broken up into discrete units and not strongly related by stating after they were done with a unit they were “finished” and “moving on” to another content idea. Although the teacher in context 1 also taught content ideas in units, the units seemed to build up knowledge over time rather than “moving on” to new separate ideas. No observations were done in the context 3, but the teacher did note that she followed their district provided book and covered the chapters in the book in that order.

Variables outside the control of the teacher also very likely influenced student achievement in the three different contexts. The students in all three contexts themselves
are very different and have different backgrounds. Although all three classrooms contained a diverse mix of typical 10th grade biology students and showed a very similar understanding of molecular genetics prior to instruction (Figure 2), the motivation of each group of students to learn may have been different and may have influenced student achievement. The classroom and school culture could also play a role in student motivation to learn and thus, influence student achievement. The method of assessment also plays a role in the level of student achievement. Student achievement in each of the contexts was higher in the interviews than on the written assessments (Figure 2). In general, the students did not explain their answers on the written assessments thoroughly, leading to lower scores. When asked for verbal explanations, the students were able to better explain their answers. The interviewer posed clarification questions which probed their understandings, leading to higher scores.

While actual student achievement in the constructs is probably closer to the interview scores, students who understand the content should also be able to sufficiently explain that understanding in writing. The timing and context of the assessments could have also affected student achievement. Students could have not taken the written assessments seriously if participation points were not awarded by the teacher or if the students did not value the purpose of the assessments. It is also unknown if the students had exams or assessments in other classes that particular day or if they just did not want or care to take the written assessment. The external factors played a role regarding student achievement; factors vary among contexts.

Because of this, it is impossible to make direct comparisons between the different classrooms but the results in the three contexts can be discussed and the differences
between the three contexts highlighted. From the data collected, context 1 supported the largest increase in student achievement in molecular genetics (Figure 2). Context 2 supported a moderate increase in student achievement, evident in the interview data (Figure 2B) but not on written assessments (Figure 2A). Context 3 also supported a moderate increase in student achievement on written assessments (Figure 2A). A limitation to this study was that no interviews were completed with students in context 3. It would have been interesting to see how these students performed in interviews and how their interview performance related to the performance of students in the other contexts.

Since the different classroom contexts are not able to be directly compared due to teacher effects; classroom, school, and district contexts; different students; and different curriculum; the students that made references to ideas explicitly mentioned in the intervention units were compared to students in the same classroom context who made no explicit references to ideas mentioned in the intervention units to determine what, if any, impact they had on student achievement. Only four students in context 2 made any explicit references to the main ideas included in the intervention units and each of the students referenced the idea in a completely different construct than the other students. Since a response by a single student cannot be compared to the rest of the classroom to determine the impact of the intervention units in that context, the impact of the units in classroom context 2 could not be determined. A larger number of students in classroom context 1 made explicit references to ideas mentioned in the intervention units, so the potential impact of the units in this context could be examined.

The only statistically significant difference between students who made two or more explicit references to protein structures and functions mentioned in the intervention
units and students who made 0-1 references was in construct B which describes the idea that genes code for protein structure and function. Students who made two or more references to protein structures and functions mentioned in the intervention units had a more sophisticated understanding of this construct in their middle interviews than students who made 0-1 references (Figure 12); this finding indicates that the examples of protein structures and functions helped students understand that genes code for protein structures and functions. Most interview students in context 1 at least understood that genes code for sub-cellular entities, such as amino acids or proteins or molecules, after instruction (Figure 4B), but the students who made two or more references to protein structures and functions were able to understand this concept more quickly than their counterparts.

The finding supports the recommendations by Roseman et al. (2006) and Duncan et al. (2009) that students should be given examples of proteins and their functions early in their molecular genetics instructional period. Unfortunately, the portion of construct C targeted in this study revolved around the idea that changes to genes change proteins (C2, Table 4.4) and not the idea that proteins are central to the functioning of cells (C1, Table 4.3). Providing students with many examples of protein structures and functions should, in theory, help students understand how proteins are central to the functioning of cells. Future studies should examine the impact of providing students with examples of protein structures and functions in relation to the new construct C1.

In three constructs, there were significant differences between students who made references to the gene, protein, cell, trait (GPCT) scaffold included in intervention unit 3. Students who made one or more references to the scaffold had a significantly more
sophisticated understanding of construct D, the idea that cells have the same DNA but express different genes (Figure 13). However, not one student made a reference to the scaffold while answering any questions related to construct D. Although the first portion of the GPCT scaffold (gene, protein, cell) is relevant to this construct as it describes how the genes code for the proteins necessary for the cell to do its function, the lack of reference to the scaffold in any answers may suggest that students do not see the scaffold as useful for this construct. The scaffold was used in intervention unit 3 to depict to students how a change to the gene can lead to genetic disorders that are observable at the trait level, such as cystic fibrosis and sickle cell anemia. The content and scaffold was presented after students had learned that cells have the same DNA but that differential gene expression makes the cells function differently. Because no students referenced the GPCT scaffold in any answers for construct D, it appears that the students who made the scaffold references understood the construct better for reasons other than the presence of the scaffold.

The second significant difference between students in a construct involved students who referenced the GPCT scaffold in construct G1, the idea that DNA varies between individuals and species. The students who made one or more references to the scaffold had a significantly more complex understanding of this construct in both the middle and post interviews than students who made no references to the scaffold (Figure 14). However, the two groups of students held different understandings prior to instruction. Although the difference prior to instruction was not statistically significant, it is likely that grouping students who referenced the GPCT scaffold versus students who did not created groups of students with inherently different abilities in this construct and
does not show evidence that the scaffold is productive for student achievement in this construct.

The revised version of this construct (G1, Table 4.8) contains a sixth level which describes the idea that the more conserved a genetic sequence is, the more important protein product it produces. This idea was not probed in any of the assessments in this study and was added as a result of the superior understanding the students had in this construct after instruction. Although the GPCT scaffold may not be helpful for student achievement up to level 5 of this construct, it would be interesting to determine if the scaffold helps students achieve the sixth level, which discusses how similar genes produce similar proteins, the first portion of the scaffold (gene, protein).

The final significant difference between students in a construct involved students who made two or more references to the GPCT scaffold versus students who made 0-1 references in construct G2, the idea that genetic changes drive evolution. The students who made two or more references to the scaffold had a significantly more complex understanding of this construct after instruction than students who made 0-1 references to the scaffold (Figure 15, post interview). The discrepancy between the two groups’ ideas prior to instruction were not statistically significant (Figure 15, pre interview), however they were a bit different. This difference became more dramatic, although still not significant, during the middle interview. Since the GPCT scaffold had not yet been discussed when the middle interview was implemented, it was not possible for the scaffold to influence student achievement during the middle interview. The difference between the groups in the final interview was statistically significant, but like construct G1, it is possible that the groupings identified students who had a better understanding of
this construct in general than showing that the scaffold was productive for student understandings of this construct.

Determining impact of the intervention units was very problematic in this study. The three classroom contexts could not be directly compared due to teacher effects; classroom, school, and district contexts; different students; and different curriculum. Grouping students who made explicit references to ideas mentioned in the intervention units versus students who did not could have merely identified groups of students who had better understandings of the constructs in general due to a variety of factors rather than just due to the influence of the intervention units. Also, identifying explicit references to the intervention units themselves was sometimes problematic as students sometimes explained that their “cancer unit” was helpful. The teacher in context 1 developed a variety of lessons on cancer, most of which she created, so any general references to a “cancer unit” or “when we talked about cancer” were not counted as references because there was no way to determine if the content was in the intervention unit dealing with cancer or a teacher-created unit on cancer. It was extremely difficult to determine if there was any actual impact of the units due to the varying classroom contexts and teacher effects.

A suggestion for a future study would be to revise the written assessments and interviews to include asking students to explain where they learned the information for each question. This modification would make the already lengthy written assessment and interviews even longer. Since it was often difficult to get students to explain why they checked an answer, it may be nearly impossible to get students to explain also where they learned that information. It may be easier to get students to explain the source of their
information verbally during an interview, but some of the 9 question interviews were already 30 minutes long and removing students for long periods of time from their normal classroom instructional time is disruptive.

For future studies, it may also be helpful to try to get all the teachers to enact all the intervention units in their entirety for more opportunities for the content in the intervention units to become useful for the students. Only four students in context 2 made any explicit references to ideas from the intervention units, so the impact of the units could not be assessed in that context due to the lack of data.

Despite there being nearly no quantitative evidence that the intervention units made any impact on student learning, some of the qualitative evidence suggests that the units were helpful. Over 70% of interviewed students in context 1 (24 of 34 students) made references to ideas explicitly included in the intervention units. The students referenced the ideas in constructs B, C, D, F, G1, G2, and H, which were the constructs specifically targeted by the intervention units. Many of the students also explained that the ideas from the units were helpful for them. Some of the students even used ideas from the intervention units to explain their answers on the written assessments. Including how they learned their information was unprompted, but some students wrote “GPCT” or “DNA is like the recipe book” when explaining why a certain answer was better than another answer. Although this study was unable to obtain much, if any, quantitative evidence that the intervention units positively impacted student learning of ideas in the progression, it is likely the units did positively impact student achievement. However, due to many more students in context 1 referencing the intervention unit ideas, it is possible that teacher effects and the way in which the units were enacted in the classroom
impact student learning more significantly than the actual intervention units by themselves.
V. Conclusion

This is the first study to fully empirically test and revise and refine all constructs of the Duncan et al. (2009) molecular genetics learning progression, and it did so in three different classroom contexts in two different schools also using novel researcher-developed intervention units targeted to six of the eight constructs and assessments targeting all of the constructs in the progression. It is also the first study to combine the two molecular genetics learning progressions (Duncan et al., 2009, Roseman et al., 2006) into a single progression.

In general, this study found that the student ideas in each construct were consistent with the levels described by the Duncan et al. (2009) progression, but several lower and intermediate ideas were identified. The empirical data obtained in the three classrooms were used to revise and refine all the constructs of the Duncan et al. (2009) progression. The student ideas were molded into productive “stepping stones” that are important conceptual shifts in student understandings and added to the progression in the forms of new levels in each construct. Two of the constructs (C, G) were split into two constructs because they both contained multiple ideas that could not be assessed at the same time since students held different understandings of the different ideas. A new hypothetical construct, I, was suggested based on ideas in the Roseman et al. (2006) progression that were not included in the Duncan et al. (2009) progression.

Since empirical validation of a progression is not a one-time study, but occurs through multiple iterative rounds of empirical studies and refinements (Shea & Duncan,
2013), this study makes a significant contribution to the field of molecular genetics education by offering eleven empirically revised and refined constructs for further studies in new classroom contexts. Additionally, the study provides assessments that target nearly all of the levels of nine of the constructs (C1, I, and level 6 of G1 being the exceptions) and curricular intervention units that target 6 of the constructs, which may be used and modified by future researchers in the field of molecular genetics education.

Looking forward, the modifications to the learning progression need to be empirically tested in additional classrooms with students in grades 5-10 to determine if students in additional contexts hold the new ideas included in the progression. A valid, reliable assessment needs to be created to probe student’s understandings of each of the constructs. Additionally, curriculum that targets instruction to the upper bounds of the progression constructs needs to be created. This study shows that students have particular difficulties integrating the three models of genetics, so curriculum and instruction should make a large effort to get students to integrate the three models and understand how they are related.

This study impacts molecular genetics education and the broader science education community by providing empirical evidence supporting the molecular genetics learning progressions and significant modifications to the progression. Since learning progressions strongly influenced the creation of the Framework, and thus, the Next Generation Science Standards, testing and modifications to progressions are important for providing empirical evidence for standards and knowing what students are capable of achieving in the classroom.
VI. References

Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS. (2013). Next generation science standards. Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS.


Appendix A

Molecular Genetics Pre/Post Test

1. Put the following terms in some sort of order or pattern:
   DNA, gene, chromosome, nucleotide/base, cell, genome

Why did you put them in that order/pattern?

Use the following paragraph to answer the next 3 questions.

_Duchenne muscular dystrophy is caused by a change in the dystrophin gene in the DNA. Normally, dystrophin anchors muscle fibers, but if the gene is changed, it cannot perform its function properly. This change results in muscle degeneration, difficulty walking, and a shortened life span._

2. The dystrophin gene is involved in anchoring muscle fibers. For each cell type (muscle and blood), place an “X” in each box where you think the dystrophin gene, mRNA, or protein are present.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Dystrophin gene</th>
<th>Dystrophin mRNA</th>
<th>Dystrophin protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle cells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood cells</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why did you place the X’s where you did?
3. Check the box next to the statement you think best explains how DNA is involved in muscle function.

☐ The genes in our DNA code for instructions for making different molecules. These molecules have different functions inside of the cells. For example, a muscle gene codes for a muscle protein that helps the muscle cells have the structure and function needed to move our legs.

☐ DNA contains genes which code for instructions for the body. These genes tell our cells how to grow, function, and develop. For example, the muscle genes in our DNA tell our muscles how to be structured and how to contract our muscle fibers so that we are able to move our legs.

☐ Our DNA has genes which code for proteins. The gene gives the order of amino acids that make up a protein. For example, a muscle gene tells what amino acids make up a muscle protein. This protein helps the muscle cells have the structure and function to move our legs.

Why did you choose this answer over the other two?

Students read more about Duchenne Muscular Dystrophy in class. They were then asked to explain how the change in the dystrophin gene leads to the muscle degeneration and difficulty walking.

4. Which student do you think best explained how the change in the gene leads to the physical effects seen with muscular dystrophy?

<table>
<thead>
<tr>
<th>Student 1</th>
<th>Student 2</th>
<th>Student 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A change to the dystrophin gene changes the amino acids in the dystrophin protein. The dystrophin protein anchors the muscle fibers, so the change to the amino acids changes the function of the protein. If the muscle fibers are not anchored correctly, they do not work properly which leads to muscle degeneration and difficulty walking.</td>
<td>Changing the dystrophin gene changes the dystrophin protein. The protein anchors the muscle fibers, so the change to the gene breaks the anchor’s function. If the muscle fibers are not anchored, they do not work properly which leads to muscle degeneration and difficulty walking.</td>
<td>The change to the dystrophin gene changed the instructions given to the cell. The dystrophin gene anchors the muscle fibers, so a change to the gene alters the anchor. If the muscle fibers are not anchored, they do not work properly which leads to muscle degeneration and difficulty walking.</td>
</tr>
</tbody>
</table>
Answer: ________

Why did you choose this answer over the other two?

Sickle cell anemia is a disease passed down through families. Red blood cells are normally shaped like a disk (#1). With sickle cell anemia, red blood cells form an abnormal crescent shape (#2). This disease is caused by a change of a single letter in the genetic code, or DNA. The letter is changed from an A to a T. This single letter change makes red blood cells turn sickle shaped.

5. The paragraph above refers to DNA as the “genetic code.”

Check the box next to the statement that you think best explains why DNA is sometimes called the “genetic code.”

☐ The genes in our DNA code for instructions for our body. These coded instructions tell our body how to function and develop amino acids. We need this code as instructions for our body to function correctly.

☐ Genes are read to determine a specific amino acid sequence that makes up a protein. Proteins do the functions in our body. Organisms use the same amino acids to make proteins and almost all use the same genetic code.

☐ Genes in our DNA code for molecules inside of our cells like amino acids and proteins. These molecules do the functions in our body. The genetic code is used by all organisms to produce molecules for its cells.

Why did you choose this answer over the other two?
The paragraph says that sickle cell anemia is caused by a single letter in the DNA being changed from an A to a T.

6. Check the box next to the statement you think best explains how the single letter change causes the red blood cells to change shape.

☐ The change in the DNA changed a protein inside of the cell. Proteins do work inside of the cell. Since the protein was changed, it can no longer make the cell round so it looks sickle shaped.

☐ The change in the DNA changed the amino acid sequence of a protein. This changed the function of the protein. The protein behaving in the new way causes the cell to look sickle shaped.

☐ The change in the DNA changes the shape of the cell. The DNA tells the cell what to look like and how to function. Because the DNA was changed, it now tells the cells to look sickle shaped.

Why did you choose this answer over the other two?

Students looked at skin and nerve cells under a microscope in class and noticed that they looked very different. They were then asked to explain why these cells looked so different.

7. Which student do you think best explained why the cells looked different?

<table>
<thead>
<tr>
<th>Student A</th>
<th>Student B</th>
<th>Student C</th>
</tr>
</thead>
<tbody>
<tr>
<td>The different cells have different functions. Nerve cells need to communicate so they have long dendrites that connect to other nerve cells in the brain. Skin cells are flat because skin needs to be flat to cover your body.</td>
<td>The different cells have different proteins inside of them. Proteins carry out the basic functions of the cell. Proteins are special to their functions. Different proteins make the cells behave and look differently.</td>
<td>The different cells have the same DNA inside of them. The cells just use different parts of the DNA to make different proteins. These different proteins make the cells look and behave differently.</td>
</tr>
</tbody>
</table>

Answer: ________
The picture to the right shows three different organisms: a fruit fly, a human girl, and a human boy.

8. Check the box next to the statement you think best explains why these three organisms look different.

- The organisms look different because their DNA is different. Humans and flies have different genes which code for the differences in the organisms. Flies have genes to grow wings and humans have genes for skin. Human boys and girls have a similar overall pattern but look different because girls have two X chromosomes while boys have an X and a Y chromosome. These differences make humans unique.

- Flies look different from humans because they have different chromosomes. Flies have 4 pairs of chromosomes while humans have 23 pairs. These chromosomes tell the body to be structured in a certain way. For example, fly DNA makes flies have wings and human DNA makes humans have skin. Human boys and girls look different because they have different human traits in their DNA.

- The organisms look different because some of their DNA is different. Humans and flies have some of the same DNA, but also have some different DNA. Humans have a similar overall pattern but look different because they have smaller differences in their DNA. These small differences make humans unique and can allow us to identify individuals.

Why did you choose this answer over the other two?
Fred and Frank are 60 year old identical twins. Fred has smoked since he was 20, but Frank does not smoke. When they were younger, people could not tell them apart. But now, Fred has a lot of wrinkles, grey hair, a cough, and has developed lung cancer. Frank has less wrinkles, grey hair, no cough, and no lung cancer.

9. Use the statements below to create an explanation for why Fred developed lung cancer and his twin brother did not. Pick one sentence from each group.

**Group A**
- A1 - Fred and Frank started out with identical DNA because they are identical twins, but Fred’s smoking changed his DNA.
- A2 - Fred and Frank had identical DNA because they are identical twins, but smoking has caused mutations to Fred’s DNA.
- A3 - Fred and Frank have identical DNA because they are identical twins. Smoking did not change Fred’s DNA, just his traits.

**Group B**
- B1 - The mutations changed genes in his cells and thus, proteins in the cells.
- B2 - Fred’s smoking caused his wrinkles, cough, and lung cancer because Frank does not have these symptoms.
- B3 - These changes to his DNA mutated his lung cells and skin cells.

**Group C**
- C1 - The smoke and tar from the cigarettes caused the changes to Fred’s body and clogged his lungs. The build up of the tar caused some of his lung cells to turn cancerous.
- C2 - Some of his lung cells now produce different proteins or more proteins. These different proteins caused the cells to turn cancerous.
- C3 - Some of his lung cells now produce mutated proteins, different proteins, or more proteins. These proteins caused the cells to turn cancerous.

Why did you choose the statements you did?
**Group A:**

**Group B:**

**Group C:**

Use the following scenario to answer the next 2 questions.

Bill works at a garden store that has a variety of different plants that have flowers. The store sells snapdragons, petunias, sunflowers, roses, daisies, and many more.

10. Which of the following statements do you best think explains why the flowering plants look different?
The plants look different because they have different plant chromosomes. The roses have rose chromosomes that tell the plant to look like a rose, while the daisies have daisy chromosomes that tell the plant to look like a daisy.

The plants look different because their DNA is different. Roses and daisies have different genes which code for the differences in the plants. Roses have genes to grow thorns and daisies have genes for long thin petals.

The plants look different because some of their DNA is different. The plants have some of the same DNA for stems and leaves, but also have some different DNA for the shapes of the flowers and leaves.

Answer: __________

Why did you choose this answer over the other two?

Different flower colors of each kind of plant are available as well. For example, the store sells both red and white-flowered snapdragons. Bill was able to cross a red-flowered snapdragon plant with a white-flowered snapdragon plant to produce plants that had pink flowers.

11. Check the box next to the statement you think best explains how Bill was able to produce pink-flowered snapdragons from crossing a white-flowered plant with a red-flowered plant.

- Snapdragons can have different versions of the same trait. The white-flowered plant has a white flower trait and the red-flowered plant has a red flower trait. When the plants were crossed, the red and white traits mixed in the new plant to make a new pink flower trait.

- Snapdragons have two copies of the flower color gene, which codes for a flower pigment protein. The gene has small changes in the DNA that make the pigment protein be white or red. The pink-flowered plant got a gene for red pigments and a gene for white pigments, so the flower is pink.

- Snapdragons have two versions of the gene for flower color pigment. Each version, or allele, is on a chromosome. The white-flowered plant gave a white flower allele to the new plant and the red-flowered plant gave a red flower allele. The genes mixed in the new plant to make the flower pink.
Bill's boss mixed up the labels for a plant fertilizer and a pesticide that is known to cause cancer in humans. Bill accidentally “fertilized” the plants with the pesticide.

12. Check the box next to the statement you think best explains what would happen to the plants after they were exposed to the pesticide.

☐ The function of the plant cells would be altered since the pesticide causes cancer in humans. The cells would stop working properly because the pesticide caused the proteins inside of the cell to change. Since proteins carry out the functions of the cell, changes to the proteins would cause the cells to not function properly.

☐ Since the pesticide causes cancer in humans, it will also mutate the plants. The pesticide could cause the plants to start growing bigger leaves or more flowers on a single plant. Each plant would react to the pesticide differently, so two red rose plants could end up looking and growing very different because of the pesticide.

☐ The plants’ DNA would be mutated since the pesticide causes cancer in humans. Changes to the DNA could cause mutations to genes. These mutations could cause the proteins to be mutated and not function properly. The mutations could also alter the expression of genes and cause some proteins to be over- or under-expressed.

Why did you choose this answer over the other two?
13. Bill consulted three scientists to see what would happen to the plants. Which scientist do you think best explained what would happen to the plants if they survived after being “fertilized” with the pesticide?

<table>
<thead>
<tr>
<th>Scientist A</th>
<th>Scientist B</th>
<th>Scientist C</th>
</tr>
</thead>
<tbody>
<tr>
<td>The plants would have increased genetic variation because their genetic material changed from the pesticide. If the mutation was beneficial, like causing the plant to make more seeds, over time, the population of plants would look more like the mutated plant because it would reproduce more.</td>
<td>The plants’ genetic material would be changed from the pesticide. These changes could change the structure and function of proteins inside the cells, causing physical changes to the plants. These could be good (bigger leaves) or bad (weaker stems) changes for the plants.</td>
<td>The plants would look and function different from normal plants. Some of the plants may have larger leaves, more flowers on the plant, or weaker stems. The plants could also grow faster or slower since they got mutated from the pesticide. Each plant would react differently.</td>
</tr>
</tbody>
</table>

Answer: __________

Why did you choose this answer over the other two?

The picture to the right shows the DNA from two different rabbits. The girl rabbit (left) has brown fur and black eyes. The boy rabbit (right) has grey fur and black eyes. The DNA from the two rabbits are shown below each rabbit. Fur color is on top and eye color is on bottom.
14. Check the box next to the statement you think best explains what the colored bars in the DNA are.

☐ The colored bars in the DNA are the traits. The girl rabbit has a grey trait and a brown trait. The rabbit has brown fur because the brown trait wins out genetically. The boy rabbit has two grey traits, so it has grey colored fur. Organisms can have different versions of traits in the DNA.

☐ The colored bars are different versions of the same gene. There are small changes in the DNA that make the fur pigment protein brown or grey. Both brown and grey pigment proteins are expressed in the girl rabbit, but it has brown fur because the brown pigment shows up more.

☐ The colored bars are alleles. The girl rabbit got one allele from its mom and one from its dad. The brown allele is dominant over the grey allele because the rabbit has brown fur. The grey allele is recessive because the boy rabbit has to have two of the grey alleles to have grey fur.

Why did you choose this answer over the other two?

15. Imagine if these rabbits mated and had baby bunnies. Check the box next to the statement you think best explains what a baby bunny would look like.

☐ There is a 50% chance the bunny will have grey fur and black eyes. There is a 25% chance the bunny will have brown fur and black eyes. There is a 25% chance the bunny will have brown fur and blue eyes. Although possible, it would be extremely rare for the bunny to have grey fur and blue eyes.

☐ There is a 75% chance the bunny would have grey fur and a 25% chance the bunny would have brown fur because there are three traits for the grey and one trait for the brown. There is a 50-50 chance the bunny would have black or blue eyes because there are two dominant black traits and two recessive blue traits in the parents.

☐ There is a 50% chance the bunny would have brown fur and a 50% chance the bunny would have grey fur. The brown is dominant over grey for fur color. There is a 75% chance the bunny would have black eyes and a 25% chance the bunny would have blue eyes because black is dominant over blue.

Why did you choose this answer over the other two?
Appendix B

Examining High School Students’ Understandings of Molecular Genetics
Full Student Interview Protocol 2011-2012

Brief Introduction (first interview):
Hi, my name is Ms./Mrs./Mr./Dr. __________. I am a researcher in science education at...

Background/Thank You: Your teacher is involved in a project that examines high school students’ understandings of molecular genetics. We are talking with students like you to help us learn more about how to improve teaching and learning in science. Thank you very much for agreeing to talk with me and thank you to your parents for allowing you to participate.

Purpose of interview/No right or wrong answers: We will be talking today about your ideas of molecular genetics. There are no right or wrong answers for any of the questions we will talk about. We are just interested in hearing your ideas. What you say on the tape will not affect your grade in your science class in any way.

Tape recording/Privacy issues: I am going to tape record the interview because I am interested in your ideas and want to be sure that I have a good record of everything you say. I’m going to ask you to run the tape recorder [have student turn tape recorder on, record a sample conversation to ensure the tape recorder is working properly, play it back to student can hear the recording, etc.]... We may share some of your ideas with teachers and researchers who are interested in students’ ideas about molecular genetics, but your name will not be connected with your ideas in any way.

Student questions: Do you have any questions about the interview?

Brief ice-breaking question: Can you tell me a little bit about yourself?
Possible follow up questions:
   How long have you attended [school name]?
   What are your favorite subjects in school? What do you like most about...?
   What do you like to do for fun?

Brief concluding comments:
[Student name], thank you for sharing your ideas. I enjoyed very much hearing your thoughts about molecular genetics. Do you have any questions you would like to ask me?

You will be doing many science activities this year related to molecular genetics. We will be videotaping you and your classmates as you learn about molecular genetics and we will also be talking with you again as you learn more about molecular genetics. We look forward to hearing how your ideas about molecular genetics change as you learn more about it and more about how molecular genetics impacts the world.

Thank you.
Interview Probe

Pre/Middle Interview - students will have a copy of their completed Pre-Test to refer to.

Post Interview - students will have a copy of their completed Post-Test to refer to.

Pre Interview
Discuss questions:
1. Put the following terms in some sort of order or pattern. (Q1)
2. Check the box next to the statement you think best explains how DNA is involved in muscle function. (Q3)
3. Which statement do you think best explains how the change in the gene leads to the physical affects seen with muscular dystrophy? (Q4)
4. Which student do you think best explained why the cells looked different? (Q7)
5. Which of the following statements do you best think explains why the flowering plants look different? (Q10)
6. Check the box next to the statement you think best explains how Bill was able to produce pink-flowered snapdragons from crossing a white-flowered plant with a red-flowered plant. (Q11)
7. Check the box next to the statement you think best explains what would happen to the plants after they were exposed to the pesticide. (Q12)
8. Which scientist do you think best explained what would happen to the plants if they survived after being “fertilized” with the pesticide? (Q13)
9. Check the box next to the statement you think best explains what a baby bunny would look like. (Q15)

For each question:
1. Can you tell me about your answer?
2. Why did you pick this answer over the others?

Middle Interview
Discuss same 9 questions as before (Q1, 3, 4, 7, 10, 11, 12, 13, 15)

For each question:
1. This was your pre-test. Do you still agree with your answer?
2. Why or why not?
3. Can you relate anything you have learned in class to this question?
4. Has anything in class helped you answer this question better?

Post Interview
Discuss same 9 questions as before (Q1, 3, 4, 7, 10, 11, 12, 13, 15)

For each question:
1. This is your post-test. Do you still agree with your answer?
2. Why or why not?
3. Can you relate anything you have learned in class to this question?
4. Has anything in class helped you answer this question better?
5. Did your thinking about this concept change over time?