Open Access Dissertations

Theses and Dissertations

Fall 2014

Diagnosing undergraduate biology students' experimental design knowledge and difficulties

Annwesa Dasgupta
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, Higher Education Commons, and the Science and Mathematics Education Commons

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Annwesa Dasgupta

Entitled
Diagnosing Biology Undergraduate Students' Experimental Design Knowledge and Difficulties

For the degree of   Doctor of Philosophy

Is approved by the final examining committee:

Nancy J. Pelaez

David Eichinger

Dennis J. Minchella

Jeffery Karpicke

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Nancy J. Pelaez

Approved by Major Professor(s): _____

Approved by: Jeffrey Lucas   12/04/2014

Head of the Department Graduate Program   Date

DIAGNOSING UNDERGRADUATE BIOLOGY STUDENTS' EXPERIMENTAL

DESIGN KNOWLEDGE AND DIFFICULTIES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Annwesa Dasgupta

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

For my Mom, Dad, Sister and Husband.

ACKNOWLEDGEMENTS

I thank my graduate committee for sharing their knowledge and providing invaluable advice throughout my doctoral journey. To Dr. Nancy Pelaez, my research advisor, thank you for being a friend, philosopher and guide. I am extremely grateful for your exemplary mentorship and for the valuable research and life lessons I learned from you. You have inspired me to be a good, hardworking researcher and teacher. Thank you for including me as part of your research group. Thank you, Dr. Anderson for your valuable guidance and support with my research and manuscripts. I have learned a lot about writing good research papers following your lead. Thank you, Dr. Eichinger for your advice with my teaching and research. The experiences I gathered as a teaching assistant with your course equipped me with teaching skills that I will continually draw upon as I embark upon my journey as a biology teacher. Dr. Minchella and Dr. Karpicke, thank you for your help and guidance honing my research goals and outcomes. Your presence as my committee members has helped me develop my research to address a wider interdisciplinary audience.

I would like to thank my research group at VIBE (Visualization In Biochemistry Education) for the constant support mentally and with research through this long, sometimes exhausting, but exciting doctoral journey. Thank you for being such wonderful colleagues and making work enjoyable. I would like to thank PIBERG

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Dasgupta, Annwesa. Ph.D., Purdue University, December 2014. Diagnosing Biology Undergraduate Students' Experimental Design Knowledge and Difficulties. Major Professor: Nancy Pelaez.

Experimental design is an important component of undergraduate biology education as it generates knowledge of biology. This dissertation addresses the challenge undergraduate educators face for assessing knowledge of experimental design in biology by examining knowledge of, and difficulties with, experimental design in the context of first-year undergraduate biology students at Purdue. The first chapter reviews several recent reports that highlight the necessity to increase understanding of the experimental research process as a core scientific ability (for e.g., AAAS, 2011; AAMC-HHMI, 2009; NRC, 2007). Despite its importance, there is limited information about what students actually learn from designing experiments. In the second chapter, the development and validation of a Rubric for Experimental Design (RED) was informed by a literature review and empirical analysis of thousands of undergraduate biology students' responses to three published assessments. The RED is a useful probe for five major areas of experimental design abilities: the variable properties of an experimental subject; the manipulated variables; measurement of outcomes; accounting for variability; and the scope of inference appropriate for experimental findings. The third chapter presents an original 'Neuron Assessment' based on a current research problem related to a disease caused by defective movement of mitochondria in neurons. This assessment provides necessary background information and figures to examine knowledge of experiments through representations and experimental design concepts. A case study method was conducted with oral interviews to investigate interactions among three factors,

conceptual knowledge (C), reasoning skills (R) and modes of representation (M). Findings indicate the usefulness of the 'Neuron Assessment' to probe knowledge and difficulties in areas characterized by RED. The fourth chapter examines evidence from the case study participants' written responses to paper and pencil tests to validate the 'Neuron Assessment' as a diagnostic tool for the RED areas. In comparison to the published assessments that formed the basis for development of RED, findings with the 'Neuron Assessment' provide strong evidence for its validity as a probe to distinguish expert and student knowledge from difficulties with experimentation concepts and representations. In summary, a mixed methods approach was used to characterize undergraduate biology students' knowledge and difficulties with experimental design. Findings from this dissertation illuminate knowledge of experimental design at the undergraduate level and open up several new avenues for improved teaching and research on how to evaluate learning about the experimental basis for understanding biological phenomena.

CHAPTER 1: INTRODUCTION

1.1 <u>Overview</u>

Scientific thinking is defined as the application of scientific methods or principles of scientific inquiry to reasoning or problem-solving situations. It involves the skills implicated in generating, testing and revising theories, and in the case of fully developed skills, reflecting on the process of knowledge acquisition and change (Koslowski, 1996; Kuhn & Franklin, 2006; Wilkening & Sodian, 2005). An important component of scientific inquiry includes designing experiments which involves evaluating evidence and making inferences in the service of processing, visualizing and interpreting explanations about a given phenomenon under investigation (Klahr, 2000; 2005a; Klahr & Dunbar, 1988). Knowledge about experimental design is an important component of biology as experiments are a way of investigating the nature of mechanisms in living systems. In its call for action, the 2011 *Vision and Change* report recommends that:

"All students need to understand the process of science and how biologists construct new knowledge by formulating hypotheses and then testing them against experimental and observational data about the living world. Studying biology means practicing the skills of posing problems, generating hypotheses, designing experiments, observing nature, testing hypotheses, interpreting and evaluating data, and determining how to follow up on the findings" (AAAS, 2011).

Thus, it is critical that undergraduate students taking biology coursework gain knowledge about identifying and designing experiments that underlie discoveries about biological phenomena. Despite the obvious importance of such knowledge in the education of biology students, surprisingly little is known about what students actually learn from designing biology experiments, compared to what they ought to learn to become competent researchers.

The need to engage biology students in experimental research has taken center stage in the past few years. There is an increasing interest in helping biology students learn about the experimental research process in general as is supported by recommendations expressed in several recent reports (AAAS, 2011; AAMC-HHMI, 2009; NRC, 2007). Undergraduate students also seem to show growing interest in biology research (Lopatto, 2003, 2008; Laursen et al., 2010; Wei and Woodin, 2011) and there has been increasing interest in course based undergraduate research experiences (CUREs) in biology (Auchincloss, 2014) which is not quite surprising, as many physical science and engineering sub-disciplines are focusing increasingly on problems related to living organisms. Increased engagement with research is justifiable as undergraduates prepare themselves to meet more rigorous academic criteria and to gain a competitive employment edge upon graduation (Laursen et al., 2010; Lopatto, 2003; 2008; Wei & Woodin, 2011). Thus an understanding of designing experiments and representing experimental results is quite evidently a core competency for undergraduate students in training as future independent researchers. But the questions that remain are: What does it mean to acquire knowledge about experiments? How can we best determine whether students are learning about experimental design including, what difficulties they have with experimental design? How do students represent their experimental design process and findings visually using for example, tables and graphs? Previous literature identifies the value of evaluating students' experimental knowledge (Kuhn and Dean, 2005; Shi et al., 2011; Sirum and Humburg, 2011).

1.2 Research Aims of This Dissertation

The research aims of this dissertation are to: (1) examine experimental design difficulties in undergraduate biology students, and (2) examine the role of assessments to improve student learning about experimental design in classrooms. Students can be taught about experimental design in the classroom but progress in their learning will require assessments that reveal knowledge of- and difficulties with experiments. Further, information about student difficulties can direct formulation of new learning outcomes in order to target areas that need specific attention by an instructor. Thus, an effective

experimental-design based course would typically carry tight alignments between learning outcomes, instructional strategies, and assessments of student knowledge. In fact in the process of course design, assessments play just as important a role as formulating learning outcomes to confirm knowledge gained in a certain area (Palomba and Banta, 1999; Pellegrino et al., 2001; Wiggins and McTighe, 1998). This dissertation examines the role of assessments in exposing students' experimental design difficulties. It further analyzes the usefulness of an original current research based assessment in detecting students' abilities with visualizations relevant to representations from experimental findings.

1.3 Dissertation Chapters

With the overarching goal of investigating students' experimental design abilities, this dissertation is comprised of three studies reported as papers, each of which describes unique approaches the exploration of deficiencies related to experimental design abilities faced by undergraduate biology students.

Chapter 2 of this dissertation summarizes Paper I which describes the development and validation of a Rubric for Experimental Design (RED) that can be used to diagnose undergraduate biology students' experimental design knowledge and difficulties. Towards achieving this goal, we conducted empirical analysis of first-year undergraduate biology students' responses to three published assessments to address the following three research questions:

1)   *What types of difficulties do undergraduate biology students have with experimental design?*

2)   *To what extent do published assessments reveal evidence of first-year undergraduate biology students' knowledge and difficulties with experimental design?*

3)   *Can the RED be usefully deployed to detect changes undergraduate students' experimental design knowledge during a first-year biology course?*

A review of the literature (Burns et al., 1985; Bullock and Ziegler, 1999; Chen and Klahr, 1999; Fuller, 2002; Kuhn and Dean, 2005; Shi et al., 2011; Sirum and Humburg, 2011) revealed the existence of a wide range of student difficulties with experimental design across multiple studies, most of which were extensively studied, with only a few that were poorly investigated. The literature survey helped us define abilities necessary for competent experimental design including: identifying a problem; generating hypotheses; planning experimental procedures with treatment, control, and outcome variables; and interpreting findings to make inferences (AAAS, 2011).

In order to examine if these problems exist among our undergraduate students, we conducted an inductive analysis of responses to three published assessments which informed the development of the RED. Five areas of difficulty with experimental design were identified: the variable properties of an experimental subject; the manipulated variables; measurement of outcomes; accounting for variability; and the scope of inference appropriate for experimental findings. The RED was also validated as an effective tool for detecting changes in undergraduate students' experimental design knowledge during instruction.

Findings from *Chapter 2* provided insight about student difficulties with knowledge of experimental design but gave no information about how students deal with visualizations, which in fact, represent a crucial part of presentation of experimental evidence (Schönborn & Anderson, 2009). *Chapter 3* of this dissertation presents Paper II which examines the potential of an original 'Neuron Assessment', which was designed based on a current research context to understand how experts and students think about experiments and visual representation of experimental evidence. Expert abilities to design an experiment and visually represent findings were first examined and used as a model to diagnose student difficulties with the same. The CRM (conceptual, reasoning and mode of representation) model (Schönborn & Anderson, 2009) was used as a guiding framework for development of the assessment. Expert and student abilities to reason with visualizations (RM) and with concepts (RC) related to experimental design were compared with the following research questions:

1. How well does the 'Neuron Assessment' reveal the nature of expert knowledge about organelle movement in neurons, and the experiments used to elucidate that knowledge?

2. How well does the 'Neuron Assessment' detect student knowledge and related difficulties with experiments to investigate organelle movement in neurons?

The experts' visualizations and knowledge of experimental concepts provided information that was used to modify our original glossary list and RED (Dasgupta et al., 2014). These were applied to examine findings from students' experimental visualizations and concepts. The 'Neuron Assessment' was found to be a good probe to distinguish expert reasoning about experiments from the performance of a typical undergraduate student. The assessment provided students with adequate information to demonstrate how they reason with visual representations (RM) and experimental design concepts (RC) and to support their ideas about investigating a current research problem.

Chapter 4 validated the 'Neuron Assessment' as a diagnostic experimental design measure by addressing the research question, *"How well does student performance on the Neuron Assessment compare with that of other assessments?"* Student participants at the undergraduate level provided written answers and diagrams for probes from three assessments which were examined for knowledge of- and difficulties with five areas of the RED (Dasgupta et al., 2014). The comparative analysis of student difficulties helped determine the usefulness of 'Neuron Assessment' to diagnose students' difficulties with the published assessments. Findings showed that students' have correct ideas with certain RED areas for a particular assessment but difficulty with others. This indicates that reasoning abilities with the RED areas are dependent on the context of the assessment. Also the 'Neuron Assessment' revealed difficulties that are not revealed in parallel by the other assessments and vice versa. Thus, different assessments should be used in combination in order to get a complete picture about student difficulties with a certain RED area.

1.4 <u>Summary</u>

This dissertation investigates how students think about experimental design and explores their knowledge and related difficulties. To that end, this research develops a Rubric for Experimental Design that showcases five major areas of experimental design difficulties. Using an original 'Neuron Assessment', visual modes of representing parts of an experiment are examined. The assessment also facilitates examination of problems in reasoning with experimental visuals as well as thinking about concepts of experimental design. Findings indicate that the 'Neuron Assessment' is a useful measure that probes for expert as well as students' experimental design ideas including visualizations like graphs.

Chapter 2 presents the RED which identifies five major areas of difficulties with knowledge of experimental design. *Chapter 3* compares expert and student abilities to reason with concepts and visualizations integral to experimental design using an original 'Neuron Assessment'. *Chapter 4* validates the 'Neuron Assessment' using a comparative analysis with other published measures of experimental design by testing its potential to diagnose knowledge and difficulties in areas targeted by RED.

<div align="center">References</div>

American Association for the Advancement of Science (AAAS) (2011) *Vision and Change in Undergraduate Biology Education*. AAAS, Washington, DC. http://visionandchange.org/finalreport

Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Dolan EL (2014). Assessment of course-based undergraduate research experiences: a meeting report. CBE-Life Sci Educ *13*, 29-40

Bullock M, Ziegler A (1999). Scientific reasoning: Developmental and individual differences. Indiv Dev from *3*, 38-54.

Burns JC, Okey JR, Wise KC (1985). Development of an integrated process skill test: TIPS II. J Res in Sci Teach *22*, 169-177.

Chen Z, Klahr D (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. Child Dev *70*, 1098-1120.

Dasgupta AP, Anderson TR, Pelaez N (2014). Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties. CBE-Life Sci Educ *13*, 265-284.

Fuller RG (Ed.). (2002). A love of discovery: Science education-The second career of Robert Karplus. Springer.

Klahr D (2000). Exploring science: The cognition and development of discovery processes. Cambridge, MA: MIT Press.

Koslowski, B (1996). Theory and evidence: The development of scientific reasoning. The MIT Press.

Kuhn D, Dean D (2005). Is developing scientific thinking all about learning to control variables? Psychol Sci *16*, 866-870.

Kuhn D, Franklin S (2006). The second decade: What develops (and how).Handbook of child psychology.

Laursen S, Hunter AB, Seymour E, Thiry H, Melton G (2010). Undergraduate Research in the Sciences: Engaging Students in Real Science. San Francisco: Jossey-Bass.

Lopatto D (2008). Exploring the benefits of undergraduate research: The SURE survey. In R. Taraban and R.L. Blanton (Eds.), Creating Effective Undergraduate Research Programs in Science. Teacher's College Press, NY. Pp. 112-132.

Lopatto D (2003). The essential features of undergraduate research. Council on Undergraduate Research Quarterly *24*, 139-142.

Palomba CA, Banta TW (1999). Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Higher and Adult Education Series. Jossey-Bass. San Francisco, CA.

Pellegrino, Chudowsky and Glaser (Eds.) (2001). Knowing what students know: The science and design of educational assessment. Washington, DC, National Academy Press.

Schönborn KJ, Anderson TR (2006). The importance of visual literacy in the education of biochemists'. BAMBED *34*, 94-102.

Shi J, Power JM, Klymkowsky MW (2011). Revealing student thinking about experimental design and the roles of control experiments. Int J Sch Teach Learn *5*, 1-16.

Sirum K, Humburg J (2011). The Experimental Design Ability Test (EDAT). Bioscene *37*, 8-16.

Wei CA, Woodin T (2011). Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE Life Sci Educ *10*, 123–131.

Wiggins G, McTighe J (1998). Understanding by design. Alexandria, VA: Association for Supervision and Curriculum Development

Wilkening F, Sodian B (2005). Scientific reasoning in young children: Introduction. Swiss J Psychol *64*, 137-139.

CHAPTER 2: DEVELOPMENT AND VALIDATION OF A RUBRIC FOR DIAGNOSING
STUDENTS' EXPERIMENTAL DESIGN KNOWLEDGE AND DIFFICULTIES

2.1 <u>Abstract</u>

It is essential to teach students about experimental design as this facilitates their deeper
understanding of how most biological knowledge was generated and gives them tools to
perform their own investigations. Despite the importance of this area, surprisingly little is
known about what students actually learn from designing biological experiments. In this
paper we describe a Rubric for Experimental Design (RED) that can be used to measure
knowledge of and to diagnose difficulties with experimental design. The development
and validation of RED was informed by a literature review and empirical analysis of
undergraduate biology students' responses to three published assessments. Five areas of
difficulty with experimental design were identified: the variable properties of an
experimental subject; the manipulated variables; measurement of outcomes; accounting
for variability; and the scope of inference appropriate for experimental findings. Our
findings revealed that some difficulties, documented some fifty years ago, still exist
among our undergraduate students, while others remain poorly investigated. The RED
shows great promise for diagnosing students' experimental design knowledge in lecture
settings, laboratory courses, research internships and Course-based Undergraduate
Research Experiences (CUREs). It also shows potential for guiding the development and
selection of assessment and instructional activities to do with experimental design.

2.2 <u>Introduction</u>

Undergraduate students are becoming increasingly engaged in biology research to
meet more rigorous academic criteria, to gain a competitive employment edge upon
graduation, or for various other reasons (Laursen *et al.,* 2010; Lopatto, 2003; 2008; Wei

and Woodin, 2011). With many physical science and engineering sub-disciplines focusing increasingly on problems related to living organisms, it is not surprising that more and more undergraduates are becoming engaged in biology research. Without biology experiments, there would be no way of investigating the nature of mechanisms in living systems; for example, how a firefly glows and how cells "know" when to divide. Designing experiments involves framing research questions to investigate observations, defining and understanding measurable variables, processing, visualizing and interpreting results.

Despite the obvious importance of experimental knowledge, and numerous calls to involve undergraduate students in authentic research experiences (Wei and Woodin, 2011), surprisingly little is known about what they actually learn from designing experiments for biological research. What has been established, though, is that experimental design is challenging for many students from elementary school to undergraduate level (Bullock and Ziegler, 1999; Burns, Okey and Wise, 1985; Chen and Klahr, 1999; Fuller, 2002; Kuhn and Dean, 2005; Shi, Power, and Klymkowsky, 2011; Sirum and Humburg, 2011). There is, therefore, increasing interest in helping biology students learn about the experimental research process in general as supported by recommendations expressed in several recent reports (NRC, 2007; AAMC-HHMI, 2009; AAAS, 2010). These reports clearly emphasize 'experimental design' as a core scientific ability. But what does it mean to acquire knowledge about experiments? How can we best determine whether students are learning about experimental design and what difficulties they might be encountering?

It is important that all undergraduate biology students experience the process of biological research as a key component of their biology curriculum. This is strongly supported by a wide range of studies in the literature that report numerous benefits to students from doing research, including a more positive attitude toward research and plans for postgraduate education in the sciences (AAAS, 2010). Most of the studies rely on rubrics (Dolan and Grady, 2010; Feldon *et al.,* 2010; Timmerman *et al.,* 2011), surveys (Kardarsh, 2000; Laursen *et al.,* 2010; Lopatto, 2004; 2007; Thiry *et al.,* 2012)

and interviews (Gutwill-Wise, 2001; Thiry *et al.,* 2012) to evaluate student learning about research. However, few of these directly measure what undergraduate students actually learned from such research experiences. There is, therefore, a gap in our knowledge in this area. In this paper we propose to address this gap through the development of a Rubric for Experimental Design (RED) that can be used to diagnose undergraduate biology students' experimental design knowledge and difficulties. Towards achieving this goal, we addressed the following three research questions:

1) *What types of difficulties do students have with experimental design?*
2) *To what extent do published assessments reveal evidence of first-year undergraduate biology students' knowledge and difficulties with experimental design?*
3) *Can a Rubric for Experimental Design (RED) be usefully deployed to detect changes in undergraduate students' experimental design knowledge during a first-year biology course?*

An overview of the research process deployed for developing and validating RED is given in Figure 2.1. To address research question 1 (RQ1) we performed a multi-step literature review (Figure 2.1A) to identify, characterize and classify known experimental design difficulties. To address research question 2 (RQ2), we deployed a process (Figure 2.1B) that identified three published assessment instruments, which were tested for their ability to detect difficulties in first-year undergraduate biology students. Data from addressing RQ1 and RQ2, namely published data about difficulties from the literature as well as data from student responses to the three published assessment instruments, were used to inform the development of RED. The RED was then tested in a pre-/post-test experimental design (Figure 2.1C) to address research question 3 (RQ3).

RQ1. Identify difficulties students have with experimental design.	RQ2. Assess students' knowledge and difficulties with experimental design.	RQ3. Use RED to detect changes in experimental design knowledge.
A1. Identify relevant literature sources for research on experimental design abilities	B1. Identify assessments with scoring materials that reveal experimental design difficulties in Table 2.1.	C1. Recruit biology majors enrolled in a first year-level undergraduate lecture course
A2. From original research reports, define features of experimental design that are problematic for students.	**B2. Use three Assessments**	**C2. Test three Assessments**
A3. Validate the list of pertinent research articles with experts.	'Shrimp Assessment'* (pre-test n=292; post-test n= 238)	Pretest / Posttest — 2010 Cohort B (n=40) / 2009 Cohort A (n=40)
	'Drug Assessment'# (pre-test n=31; post-test n= 205)	2012 Cohort D (n=31) / 2011 Cohort C (n=40)
	'Bird Assessment' ** (pre-test n=220; post-test n= 187)	2011 Cohort C (n=40) / 2010 Cohort B (n=40)
A4. Characterize known difficulties with experimental design in Table 2.1 as *Partially Established* or *Established.*	B3. Inductive analysis of 1173 responses →experimental design difficulty themes→ RED.	C3. Use RED to detect differences in 'difficulties' versus 'correct' experimental design knowledge.

Figure 2.1: Process for developing and validating the Rubric for Experimental Design (RED)
The process for developing and validating the Rubric for Experimental Design (RED) involved (A) A systematic review of the literature to identify experimental design difficulties documented by research, (B) Testing three published assessments by looking at more than 1100 responses to see how well they probe for difficulties consistent with research on experimental design difficulties from the literature, and (C) Recruiting four cohorts of students to take the assessments to develop a Rubric for Experimental Design (RED) based on their responses to published assessments collected before and after an introductory biology course. The assessments are used with permission from # SRI International, and The College Board *2006 and **2009.

2.3 Literature Review

To find out about the difficulties undergraduate biology students have with experimental design (RQ1), as per Figure 2.1A, our first step was to review the literature. This would also enable us to define the abilities necessary for competent experimental design, including identifying a problem; generating hypotheses; planning experimental procedures with treatment, control and outcome variables; and interpreting findings to make inferences (AAAS, 2010). For the literature review, we first tracked down original research from two reports from the National Academies (Duschl *et al.*, 2007; Singer *et al.*, 2005). This helped us to identify key peer-reviewed journals from disciplines ranging

from psychology and cognition to discipline-based education research journals, including those used by cell biologists, physiologists, and ecologists. Original research on difficulties was also found in articles from peer-reviewed journals in the areas of teacher education and undergraduate education (such as *Journal of College Science Teaching* and *American Biology Teacher*) and in dissertations. We did not use any secondary sources except to identify references to primary sources we might have missed. Although our main interest is in undergraduate difficulties, we included studies from child development because of the possibility that our undergraduate students might still demonstrate difficulties that have been documented by research studies on experimental design abilities with children. Within each area we identified research articles that address student difficulties or abilities related to one or more aspect of experimental design. This process helped us compile an initial list of findings from research, which was reviewed by a scientist, a cognitive scientist, a science teacher educator, and checked against references presented at a Symposium on Psychological Sciences, *Psychology of Science: Implicit and Explicit Processes* (Conference on the Psychology of Science, 2010).

Some difficulties with experimental design had rich descriptions and solid evidence, while for others we found limited evidence. For this research study, we elaborated on Grayson *et al.*'s (2001) framework to characterize and classify these experimental design difficulties as follows (Figure 2.1A4). Difficulties were classified as *Established* if they met the following criteria: (a) identified in at least three studies, (b) found in two or more different populations, (c) showed evidence that the difficulty was more than just the direct result of a single assessment, and (d) appeared with reasonable prevalence in data that supported a stable description of the difficulty. In contrast, difficulties were classified as *Partially Established* if they had been: (a) documented only in one or two studies, and (b) could have been the result of a single assessment or the way those students were taught. With limited evidence, a *Partially Established* difficulty merits further research. But with increasing triangulation of data and multiple observations in different contexts it was considered that the identified difficulty was an authentic part of student thinking rather than a function of how a particular textbook presented material, how a particular teacher taught, or from the nature of a particular

question. By classifying the difficulties in this manner, we would know which *Partially Established* and *Established* difficulties we could confidently use to inform the development of the rubric. Any remediation of such difficulties would, therefore, be based on sound knowledge of the nature of the difficulty. Of course some of the difficulties were later classified at a higher level based on our own data generated while addressing RQ1.

As summarized in Table 2.1, we found that most of the reported difficulties with experimental design could be classified as *Established* while only a few met our criteria of *Partially Established* due to limited evidence. The difficulties we found fell into five categories as listed in Table 2.1: the experimental subject itself (Difficulty I), variables (Difficulty II, A-F), measures of experimental outcomes (Difficulty III), dealing with variability (Difficulty IV, A-E), and interpreting experimental conclusions (Difficulty V, A-B). As shown in Table 2.1, difficulties were found across different populations of students at multiple educational levels, including elementary, middle and high school, undergraduates who were not science majors, and undergraduate science students.

Table 2.1: Experimental design difficulties classified on the 4-level framework and how they relate to what three published assessments measure.

[a] A review of the literature revealed that student difficulties with experimental design knowledge could be organized into five categories I-V. For definitions of the terms under I-V refer to 'Glossary of Terms' in Supplementary Information page 20;

[b] Based on the four-level framework (Grayson *et al.,* 2001), "Level" refers to how much insight there is about a particular difficulty. Difficulties found across different populations of students at multiple educational levels are classified as "*Established*"; others that require further research were classified as "*Partially established*".

[c] U: Undergraduate Students; UN: Undergraduate Science Non-Majors; UB: Undergraduate Biology Students; ES: Elementary Students; MS: Middle School Students; HS: High School Students.

[d] x's represent cases where scoring materials from the publishers claim the assessment measures knowledge consistent with the difficulty documented by past research.

	Difficulty[a]	Level[b]	Demographic Population[c]	Published Assessments[d]		
				Shrimp	Drug	Bird
I.	**Identifying the experimental subject** (Salangam, 2007)	Partially Established	UN	x	x	x
II.	**Variables: A variable property of an experimental subject**					
A.	Categorical (Discrete) variable (Picone *et al.*, 2007)	Partially Established	UN			
B.	Quantitative (Continuous) variable (Colon-Berlingeri and Burrowes, 2011; Gormally *et al.,* 2012; Harker, 2009; Hiebert, 2007; Picone *et al.*, 2007)	Established	UB			
C.	Treatment (Independent) variable (Beck and Blumer, 2012; Burns et al., 1985; D'Costa and Schlueter, 2013; Dolan and Grady, 2010; Griffith, 2007; Harker, 2009; Hiebert, 2007; Koehler, 1994; Libarkin and Ording, 2012; Picone *et al.*, 2007; Salangam, 2007; Tobin and Capie, 1982)	Established	MS; HS; UN; UB	x	x	x
D.	Outcome (Dependent) variable (Beck and Blumer, 2012; Burns et al., 1985; D'Costa and Schlueter, 2013; Dolan and Grady, 2010; Griffith, 2007; Harker, 2009; Koehler, 1994; Libarkin and Ording, 2012; Picone *et al.*, 2007; Salangam, 2007; Tobin and Capie, 1982)	Established	MS; UN; UB	x	x	
E.	Control (Comparison) group (Bullock and Ziegler, 1999; D'Costa and Schlueter, 2013; Dolan and Grady, 2010; Gormally *et al.*, 2012; Harker, 2009; Hiebert, 2007; Shi *et al.*, 2010).	Established	ES; MS; U		x	
F.	Combinatorial reasoning (Karplus by Fuller, 2002; Lawson and Snitgen, 1982; Lawson *et al*., 2000; Tobin and Capie, 1981a)	Established	MS; HS; U	x	x	x

Difficulty[a]	Level[b]	Demographic Population[c]	Published Assessments[d]		
			Shrimp	Drug	Bird
III. **Measurement of results** (Dolan and Grady, 2010; Harker, 2009; Hiebert, 2007; Salangam, 2007; Tobin and Capie, 1982)	Established	MS; UB	x	x	x
IV. **How to deal with variability:**					
A. Recognition of natural variation within a biological sample (Kanari and Millar, 2004; Picone *et al.*, 2007)	Established	MS; UB		x	
B. Random (representative) sample (Colon-Berlingeri and Burrowes, 2011; Gormally *et al.*, 2012; Metz, 2008)	Established	UB		x	
C. Randomization of treatments (Colon-Berlingeri and Burrowes, 2011; Gormally *et al.*, 2012; Hiebert, 2007)	Established	UB	x	x	x
D. Replication of treatments (Harker, 2009; Kanari and Millar, 2004)	Established	MS; UB	x	x	x
E. Reducing effect of unrelated variables (Chen and Klahr, 1999; D'Costa and Schlueter, 2013; Kuhn and Dean, 2005; Tobin and Capie, 1982)	Established	ES; MS; UB	x	x	x
V. **Interpretation of experimental conclusions**					
A. Scope of inference /generalizability of results (Chen and Klahr, 1999; Colon-Berlingeri and Burrowes, 2011; Lawson *et al*., 2000; Metz, 2008; Tobin and Capie, 1982)	Established	ES; MS; U	x	x	x
B. Cause and effect conclusions (Dolan and Grady, 2010; Griffith, 2007; Gormally *et al.,* 2012; Grunwald and Hartman, 2010; Harker, 2009; Hiebert, 2007; Klahr *et al.,* 1993; Kuhn and Pearsall 2000; Kuhn, Schauble and Garcia-Mila, 1992; Libarkin and Ording, 2012; Metz, 2008; Park and Pak, 1996; Roth *et al.*, 1998; Schauble, 1990; Schauble, 1996).	Established	ES; MS; U	x	x	

A surprising finding by Salangam (2007) is that some students do not know how to identify the experimental subject (Difficulty I). This difficulty is classified as *Partially Established* because it was found in only one quasi-experimental study with

undergraduate students who were not science majors. Further research is needed to establish to what extent this difficulty is found across different populations of students.

Thinking about and working with different variables presents students with a variety of difficulties (Table 2.1, Difficulty II, A-F). Elementary school students are known to struggle with experimental controls, and they are more competent in recognizing than designing such controls (Bullock and Ziegler, 1999). Manipulation of experimental variables is difficult for middle and high school students. This fact has been known for 50 years since Karplus first demonstrated that students have problems with formal operational reasoning patterns like combinatorial reasoning, or the simultaneous manipulation of two independent variables in a study (Fuller, 2002). Middle and high school students also have trouble identifying a treatment, outcome, and control variable (Burns *et al.*, 1985; Dolan and Grady, 2010). Gormally *et al.* (2012) recently reported that biology undergraduate students in a general education course still have difficulties with quantitative variables. Another problem undergraduate students have with treatment and outcome variables is inappropriately associating these variables in constructing a testable hypothesis (Beck and Blumer, 2012; D'Costa and Schlueter, 2013; Griffith, 2007; Harker, 2009; Libarkin and Ording, 2012; Salangam, 2007). These problems, associating treatment and outcome variables, have also been reported among undergraduates outside of biology, such as in psychology (Koehler, 1994). Even undergraduate biology majors have trouble understanding quantitative variable concepts like probability distributions, statistical p-values, and regression analysis (Colon-Berlingeri and Burrowes, 2011; Harker, 2009; Hiebert, 2007). They also have problems creating graphs from raw quantitative data (Picone *et al.,* 2007), and with treatment and outcomes (D'Costa and Schlueter, 2013; Picone *et al.,* 2007) and control variables (D'Costa and Schlueter, 2013; Hiebert, 2007; Harker, 2009; Shi *et al.,* 2010). While we classified these as *Established Difficulties*, we found only one study that exposed difficulties science non-majors' have graphically representing categorical variable data (Table 2.1, Difficulty II A). This single report about categorical variable difficulties (Picone *et al.,* 2007) was classified as *Partially Established* because further investigations are required to establish whether the difficulty is limited to graphs or if students also

struggle with the concept of categorical variables in general. Moreover, research is needed to test for this difficulty with other relevant populations such as biology majors.

Several studies have established that from middle school to biology undergraduate levels, students often fail to state their findings accurately in a way that relates to the actual measures used in an experiment (Difficulty III). Making decisions about what variables to measure at various stages of an experiment is also poorly understood by many students (Dolan and Grady, 2010; Harker, 2009; Hiebert, 2007; Tobin and Capie, 1982). Biology students who are not science majors have difficulty distinguishing between the relevant and unrelated variables that they need to measure to address a given experimental goal (Salangam, 2007).

Student difficulties with natural variability have been well documented in multiple studies that examined students doing experiments (Table 2.1, Difficulty IV). For example, some elementary and middle grade students do not understand how variability might be controlled by reducing effects of unrelated variables (Difficulty IV E) (Chen and Klahr, 1999; Kuhn and Dean, 2005), while middle school students have trouble interpreting findings when faced with natural variation (Difficulty IV A) (Kanari and Millar, 2004). Dealing with natural variation (Difficulty IVA) is also a difficult task for undergraduate biology majors and non-majors (Picone *et al.,* 2007). Biology students have difficulty reducing the effect of unrelated variables in their experiments (Difficulty IV E) (D'Costa and Schlueter, 2013). Few undergraduate students know that random assignment of treatments to samples of experimental subjects (Difficulty IV C) provides a way to measure and minimize the effect of natural variation in samples (Hiebert, 2007). Studies show that some middle school students fail to see the need to replicate treatments as a way to deal with variability (Difficulty IV D) (Kanari and Millar, 2004), while biology undergraduates show a similar problem (Harker, 2009). Undergraduate biology students also have trouble with randomization of treatments (Difficulty IV C) and the idea of having a representative sample of experimental subjects (Difficulty IV B) (Gormally *et al.,* 2012). Colon-Berlingeri and Burrowes (2011) and Metz (2008) demonstrated that biology undergraduates have difficulty summarizing trends from data

with probability distributions, and they fail to use distributions to provide information about variation and representativeness of an experimental sample (Difficulty IV B). In summary, students of all ages clearly struggle to deal with variability in an experiment.

Problems with interpreting experimental findings are another well-documented difficulty. Students from elementary (Chen and Klahr, 1999), middle school (Tobin and Capie, 1982) and undergraduate levels (Lawson *et al.*, 2000; Tobin and Capie 1981a) struggle with estimating the extent of inferences made from experimental findings (Table 2.1; Difficulty V). Another extensively reported issue (Difficulty V B) is making claims about cause and effect relationships in experiments. This problem is prevalent among students from elementary to the undergraduate level (Libarkin and Ording, 2012; Schauble, 1996).

It is surprising to note that experimental design difficulties have met our *Established* or *Partially Established* criteria as long as 50 years ago, and yet these difficulties persist with a range of students from elementary school to undergraduate levels. Undergraduate biology instructors may be unaware that these well-documented difficulties may be a challenge for their own students. Using the previously identified difficulties, we set out to find tools for diagnosing these problems in our own undergraduate biology students, because without explicit information about their problems, we would not be able to intervene with appropriate guidance.

2.4 Methods

2.4.1 Study Design

Four cohorts of approximately 300 undergraduate biology majors participated in the study at a research university in the Midwest region of the United States, across four semesters in three consecutive years (2009-2012). These students were enrolled in a first year-level lecture course on Development, Structure, and Function of Organisms. As described by Clase, Gundlach and Pelaez (2010), according to the expected outcomes for

this course, students would learn about development, structure, and function of organisms based on information from biological research such as experiments.

Many published assessment instruments for experimental design were tested of which three were selected, based on the claims of the authors (College Board 2006, 2009; SRI international, 2003) that they probe the difficulties consistent with previous literature (see Figure 2.1). These three were used as pre- and post-tests on our undergraduate biology student sample (Figure 2.1B), at the beginning and end of the semester during three consecutive years (Figure 2.1C). All assessments had been professionally validated (College Board 2006, 2009; SRI international, 2003) for use with high school students as measures for experimental design knowledge in areas I-V (Table 2.1). As a result of using each assessment with two different cohorts, we developed the RED to summarize areas where students consistently demonstrate difficulties with experimental design. Thus, this study examined whether these assessments also provide useful diagnostic information about college students.

2.4.2 Addressing Research Question 1: What types of difficulties do undergraduate biology students have with experimental design?

This question was addressed under the above literature review section. Studies of experimental design difficulties with children were included because the same types of difficulties were also reported in studies with undergraduate students (Table 2.1).

2.4.3 Addressing Research Question 2: To what extent do published assessments reveal evidence of first-year undergraduate biology students' knowledge and difficulties with experimental design?

Motivation for Selection of Assessments. For this study, three published assessments were used as diagnostic questions. With a list of important experimental design difficulties as the target (Table 2.1), the first criterion for selecting such assessments was whether publishers claim that a test probes for the difficulties documented in the literature. The published assessments that probe for experimental knowledge relevant to each category of difficulty (Table 2.1, I-V) used in this study will

be referred to as the 'Shrimp,' the 'Drug' and the 'Bird' assessments, published by the College Board (2006), SRI International (2003) and the College Board (2009) respectively (Figure 2.1).

For the 'Shrimp Assessment', students had to propose an experiment to combine nutrients and salt levels to find their effect on the growth of tiger shrimp. The 'Drug Assessment' asked students to design an experiment with appropriate patients to test a new drug for reducing high blood pressure. The 'Bird Assessment' was framed around the design of an experiment to treat pesticide granules with two different colors and patterns to find out which of the two treatments the various bird species (blackbirds, zebrafinches, and geese) will avoid eating and if there is a difference for males and females. The actual probes and scoring guidelines are included with permission and a URL for the original source of each assessment as Supplementary Information. In the Results section, we compare features of experimental design probed by each assessment to the difficulties identified from a review of the literature (Table 2.1).

The 'Shrimp Assessment.' According to the published source, an assessment from the 2006 College Board AP Statistics test (henceforth 'Shrimp Assessment') is useful for evaluating abilities to: *"(1) identify the treatments in a biological experiment; (2) present a completely randomized design with replications to address the research question of interest; (3) describe the benefit of limiting sources of variability; and (4) describe the limitations to the scope of inference for the biologist"* (The College Board, 2006, Scoring Guidelines p. 16). As per Table 2.1, this assessment measures knowledge about the experimental subject (Difficulty I), treatment or independent variables (Difficulty II C, II D, II F), measurement of results (Difficulty III) how to deal with variability with randomization and replication of treatments (Difficulty IV C, IV D), and by selecting one shrimp species as experimental subject (Difficulty IV E), and interpretation of experimental findings (Difficulty V). Thus clearly this assessment was appropriate for the present study as it is claimed to cover a wide range of difficulties. In the present study we aimed to confirm this claim and to establish whether other difficulties were revealed by this assessment.

The 'Drug Assessment.' The 'Drug Assessment' from an online database, Performance Assessment Links in Science (SRI international, 2003), asks students to design a controlled study to develop a new experimental drug for blood pressure patients. This assessment was developed by the New York State Education Department to test for experimental design abilities in a medical context. According to the authors, this assessment is designed to measure experimental reasoning abilities like "*(1) stating hypothesis, (2) organizing experimental groups, (3) selecting participants in an experiment, (3) measurement of experimental results, and (4) drawing cause and effect claims from experimental findings.*" Based on these claims, this assessment probes for various difficulties listed in Table 2.1. The assessment asks students to propose a hypothesis by associating appropriate treatment and outcome variables (Difficulty II C, II D), organize appropriate treatment and control groups (Difficulty I, II C, II D), propose measurable outcomes (Difficulty III), and account for variability sourced from unrelated variables through randomization and replication of treatments (Difficulty IV A - E). In addition, the assessment probes for cause and effect claims (Difficulty V) by which the authors make reference to interpretation of findings (Difficulty V) as well as the need to closely match the groups carrying treatment and control variables (Difficulty II C, II E).

The 'Bird Assessment.' A modification of the 2009 AP® Statistics assessment was framed around the design of an experiment to study feeding habits of various bird species (henceforth 'Bird Assessment'). This assessment was centered on statistical abilities for experimental design. According to the authors, the primary goals of this assessment were to assess students' ability to "*(1) describe assignment of experimental units to treatments in a block design and (2) provide ways to increase the power of an experiment.*" These goals align with some of the Table 2.1 difficulties because groups of experimental subjects to be tested should be considered based on a variable property appropriate for the goal of an investigation (Difficulty I), and a treatment was to be applied to groups of birds as experimental subjects (Difficulty II C, II F). Power of an experiment can be increased by replication of treatment conditions (Difficulty IV D) and also by reducing influence of the unrelated variables (Difficulty IV E). Finally, a good

experiment would focus on appropriate measurements (Difficulty III) for the proposed interpretation of the experimental findings (Difficulty V).

Based on Table 2.1, one would expect to find the same *Established* or *Partially Established* difficulties identified in previous research in the responses from undergraduate students to the assessments. In addition, one would expect data that will permit the above *Partially Established* difficulties to be re-classified as *Established*. To test these predictions, the three assessments were administered to diagnose difficulties with experimental design among our own undergraduate student population.

To identify difficulties undergraduate biology students have with experimental design, more than 1100 responses to three assessments undergraduate biology student were examined and coded for their correct ideas or difficulties with experimental design. A range of responses gathered both before and after a first-year biology course included more than 500 responses to the 'Shrimp Assessment,' more than 400 responses to the 'Bird Assessment,' and 236 responses to the 'Drug Assessment,' as illustrated in Figure 2.1B. Both inductive analysis of student responses to the assessments and the scoring materials from the publisher were used to characterize both the correct ideas and the difficulties expected from the literature review in Table 2.1.

2.4.4 Development of the Rubric for Experimental Design (RED).

Using both the published difficulties in Table 2.1 and all responses to each published assessment from volunteers collected over a period of three years, two coders started examining and coding for the students' difficulties. The coders had both completed graduate coursework in education research and both were experienced lab scientists who are familiar with experimental design. Each coder coded responses independently and then came together to discuss codes to resolve any coding discrepancies. Coding was done blindly as to whether a particular response was from pre- or post-instruction. First, qualitative analysis was performed on responses to the 'Shrimp Assessment' using inductive coding to detect recurrent mistakes. The analyses involved discriminating accurate and flawed responses and assigning unique codes for each type of

error. During inductive analysis, difficulties and accurate responses were read a number of times in order to discover similarities and emerging themes. Themes with similar meaning were coded together and grouped into a particular category (Table 2). Any discrepancy with categorizing responses either under existing codes or creating new ones was discussed until agreement was reached. This method resulted in development of RED as a rubric that represents all the difficulty themes under a particular category.

2.4.5 Addressing Research Question 3: Can a Rubric for Experimental Design (RED) be usefully deployed to detect changes in undergraduate students' experimental design knowledge during a first-year biology course?

2.4.5.1 Administering the Assessments

All assessments were administered, both pre- and post-instruction, via online Qualtrics® survey software and open-ended responses were collected as part of a regular homework assignment at the beginning and end of the semester each year. Students were given up to 10 points for providing their own ideas and thoughtfully written responses to the questions without consulting other sources. The survey took up to 30 minutes of their time. Most students enjoyed knowing that their ideas would be used to help improve instruction for students like them and they appreciated the opportunity to get points for explaining their own ideas. Different assessments were used for pre- and post-tests during a given semester to control for the same students absorbing knowledge by remembering and discussing what was asked when they attempted the test at the beginning of the course (Figure 2.1C).

2.4.5.2 Analysis of Responses

Student performance across four cohorts was examined to test our null hypothesis that the 'Shrimp', 'Drug' or 'Bird Assessment' is *not* appropriate for showing differences in the proportion of students with correct ideas or difficulties in an area of experimental design knowledge at the beginning compared with the end of a semester. Our alternate hypothesis is that the 'Shrimp', 'Drug' or 'Bird Assessment' *is* appropriate for showing

differences in the proportion of student with correct ideas or difficulties in an area of experimental design knowledge at the beginning compared with the end of a semester. To test our hypothesis, we sampled responses using a random sampling approach and examined student responses for experimental design difficulties. In spite of groups being of different sizes across four cohorts A-D, during random sampling each response had an equal probability of selection for all students (Kish, 1965). Pre and post responses were de-identified and blind coded to control for bias during analysis. Using the RED, sampled responses were coded independently by the first author once two independent coders achieved a high degree of inter-rater reliability, as reported below. As responses were coded, the sample size was gradually increased until student difficulties appeared in a consistent manner and finally reached saturation. In this study, saturation was found with a sample of 40 responses per assessment. This means that after analyzing 40 responses, we recurrently found all difficulties listed in Table 2.2 and further did not detect any new difficulties.

All responses to a particular assessment were collected as a pretest at the beginning of the semester and then all responses to the assessment were collected from a different class as a posttest at the end of the semester (Figure 2.1C). Each pre- and post-test response was assigned an individual random number using the random number generator function within MS Excel. Then, for each assessment, the 40 lowest random numbers were selected from the pre-test and 40 more were added from the post-test responses. This sampling process yielded an adequate uniform sample size to focus on the research questions and yet was manageable for classifying experimental abilities given the qualitative nature of our coding approach. A random sample of the responses was used to reduce bias during coding and to allow for representation of the overall population (Rubin, 1973). When the same assessment was used at the beginning of the semester with one class and at the end of the semester with another class we would expect to see a difference in results with students who have not taken this course (at the beginning) compared with those who have completed the course (at the end of course) provided these assessments are useful to characterize learning about experiments in this course.

To find out if each published assessment could detect changes in student knowledge as a result of course participation, Fisher's exact test was applied to detect differences in correct knowledge and difficulties with experimental design knowledge at the beginning and at the end of a semester. The Fisher's exact test is appropriate when dealing with independent samples (Ramsey and Schafer, 2012). For this study, responses from one group of students before the course were compared with responses from a different population at the end of another semester using the same assessment. In other words, data collected from these two independent random samples produced results that fell into one of two mutually exclusive classes; to determine whether they differed, we compared the proportion with answers that were correct or showed a difficulty. Further, in order to characterize how well each assessment probes for experimental design knowledge with each of the three assessments, we calculated the % of students that expressed correct knowledge and difficulties for each broad area across responses to three assessments at the beginning and at the end of a semester.

2.4.5.3 Coding of RED Areas of Difficulty

Each response was assessed for evidence of difficulties. If a problem was found based on the RED, it was coded as a difficulty under the corresponding broad area (Table 2.2). For example, a difficulty with *randomization* in the 'Shrimp Assessment' was noted under 'Randomized design of an experiment' (Table 2.2, Area of Difficulty 4-d, e, f). For each of the five big areas, if the student showed evidence of any difficulty with underlying components, that response was coded under 'difficulty' for that big area. A difficulty with any one component under area *accounting for variability* would count as a difficulty for this overall area.

Second, if we found no difficulty, we looked for evidence that shows clear understanding. Finally, if a response did not show evidence (correct or flawed) about a certain broad area, it was listed as 'lack of evidence' for that area. For example, a 'Shrimp Assessment' response stating 'measure effect of nutrients/salinity on shrimp' was considered as lack of evidence (LOE) for area *measurement of outcome* because no

indication for what to measure (shrimp growth) was characterized by the phrase, "measure effect."

At the same time as difficulties were identified, a corresponding statement was written to describe knowledge that represents correct understanding of each area based on clear definitions of key experimental design concepts (Refer to 'Glossary of Terms' in Supplementary Information). For the five areas, this was done by reviewing the literature for statements of correct knowledge. Accurate statements were validated with expert faculty and graduate students over a three year period using an iterative process until consensus was reached. The experts included a biologist who was head of undergraduate programs, a biochemist, four science education graduate students, and members of a faculty learning community that involved faculty members from biology and statistics departments. Examples of data to illustrate typical difficulties for each correct idea are presented below as well as in Supplementary Information (Tables SI 1-6). The corresponding accurate statements are listed in Table 2.2 under *"Propositional Statements/Completely Correct Ideas."*

2.4.5.4 Inter-rater Reliability

Two raters (first author and another graduate student) coded each response in terms of five areas in RED (Table 2.2). In order to initially familiarize the second coder, response examples with correct and flawed responses to each assessment were used to carefully understand the RED and further apply it to characterize student responses (See Supplementary Information Tables SI 1-3). Once 100% agreement with the RED was reached for coding the sample, the coders separated to code independently. A sample of 10 responses for three assessments each (30 responses total) was coded using the analysis approach described. To examine reliability of coding across raters, overall area codes were compared. In other words, if rater A coded a response showing difficulty for area *measurement of outcome*, we checked whether rater B also coded the response as 'difficulty' or 'correct' under *measurement of outcome*. To statistically estimate the degree of agreement as per five areas, a *Cohen's kappa* value was coded for each area on each assessment individually (Cohen, 1960). Cohen's kappa is considered a better

measure of inter-rater agreement than the simple percent agreement calculation because it adjusts for the amount of agreement due to chance. A resulting Cohen's *kappa* value of κ = 0.68 would indicate substantial agreement (Landis and Koch, 1977), meaning that with careful definition of the coding protocol and well-trained coders, responses to each assessment could be reliably coded and scored.

2.5 Findings

In addressing RQ1, the literature review (Table 2.1) revealed that most authors had identified several major categories of difficulty, all of which were classified by us as *Established*, except for two difficulties, which had limited available evidence and were classified as *Partially Established*. It is important to note, though, that most authors failed to present data that allowed them to unpack or characterize each difficulty category into sub-categories that would be more useful to instructors. In addressing RQ2, our qualitative data from the undergraduate biology students' responses to the three selected assessment instruments allowed us to significantly extend the literature knowledge to include multiple sub-categories of difficulty allowing us to develop the RED. To ensure that RED would be useful to characterize both correct and flawed responses, we pooled data from both pre- and post-tests which made it more likely to cover the full range of qualities of understanding about experimental design. In addition, to optimize confidence in our data to inform RED, we only used *Established* and *Partially Established* difficulties based on the literature review (RQ1) that included only primary research reports.

In this section, for reader convenience, we first present and describe the RED, and thereafter we present the detailed data used to inform the development and validation of this rubric.

2.5.1 The Rubric of Experimental Design (RED)

To understand, *what types of difficulties undergraduate biology students have with experimental design,* besides the data from the literature review (RQ1), all answers to three assessments were examined to identify difficulties documented in the literature as well as other flawed responses using an iterative process over a period of three years. This process led to the development of the RED (Table 2.2) with five major categories of student difficulties with experimental design as themes: *(1) variable property of an experimental subject; (2) manipulation of variables; (3) measurement of outcome; (4) accounting for variability* and *(5) scope of inference based on the findings.* These five categories form the basic framework for the RED, with multiple sub-categories of difficulty under each major category (Table 2.2). When the RED was tested for inter-rater reliability as described above, the average *kappa* value obtained was 0.9 (See Supplementary Information Tables SI 7-9 for detailed calculations), assuring high inter-coder reliability (Landis and Koch, 1977). Perhaps not surprisingly, when the RED was used as a guide to characterize and distinguish responses with difficulties from accurate responses, those with difficulties were consistent with low scores according to the scoring guidelines published by authors of the assessments (See *Scoring Guidelines* in Supplementary Information). In the sections below we present (Table 2.3) and discuss the detailed data that supported the formulation of the RED.

Table 2.2: Rubric for Experimental Design (RED).

Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
(1) Variable Property of an Experimental Subject	**Experimental subject or units:** The individuals to which the specific variable treatment or experimental condition is applied. An experimental subject has a variable property. A variable is a certain property of an experimental subject that can be measured and that has more than one condition.	**a.** An experimental subject was considered to be a variable. **b.** Groups of experimental subject were considered based on a property *that diverges* from the subjects that were the target for the stated investigation or claim to be tested. **c.** Variable property of experimental subject considered is not consistent throughout a proposed experiment.
(2) Manipulation of Variables	**Testable hypothesis:** A hypothesis is a testable statement that carries a predicted association between a treatment and outcome variable. (Ruxton and Colegrave, 2006).	**a.** Only the treatment and/or outcome variable is present in the hypothesis statement. **b.** Hypothesis does not clearly indicate the expected outcome to be measured from a proposed experiment.
	Treatment group: A treatment group of experimental subjects or units is exposed to experimental conditions that vary in a specific way (Holmes, Moody and Dine, 2011).	**c.** Haphazard assignment of treatments to experimental units in a manner inappropriate for the goal of an experiment. **d.** Treatment conditions proposed are unsuitable physiologically for the experimental subject or inappropriate according to the goal of an investigation.
	Combinatorial reasoning: In experimental scenarios when two or more treatment (independent) variables are present simultaneously, all combined manipulations of both together are examined to observe combinatorial effects on an outcome.	**e.** Independent variables are haphazardly applied, in scenarios when the combined effects of two independent variables are to be tested simultaneously. **f.** Combining treatments in scenarios where the effect of two different treatments are to be determined individually
	Controlling outside variables: The control and treatment groups are required to be matched as closely as possible to equally reduce the effect of lurking variables on both groups (Holmes, Moody and Dine, 2011).	**g.** Variables unrelated to the research question (often showing a prior knowledge bias) are mismatched across treatment and control groups.
	Control group: A control group of experimental subjects or units, for comparison purposes, measures natural behavior under a normal condition instead of exposing them to experimental treatment conditions. Parameters other than the treatment variables are identical for both the treatment	**h.** The control group does not provide natural behavior conditions because absence of the variable being manipulated in the treatment group, results in conditions unsuitable for the experimental subject. **i.** Control group treatment conditions are inappropriate for the stated hypothesis or experiment goal.

Table 2.2: Rubric for Experimental Design (RED).

Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
	and control conditions. (Gill and Walsh, 2010; Holmes, Moody and Dine, 2011).	**j.** Experimental subjects carrying obvious differences are assigned to treatment vs. control group.

(3) Measurement of Outcome	Treatment and outcome variables should match up with proposed measurements or outcome can be categorical and/or quantitative variables treatments
	    A categorical variable sorts values into distinct categories.
	    A quantitative or continuous variable answers a "how many?" type question and usually would yield quantitative responses.
		a. No coherent relationship between a treatment and outcome variable is mentioned.
		b. The treatment and outcome variables are reversed.
		
		c. An outcome variable that is quantitative is treated as a categorical variable.

	Outcome group: The experimental subject carries a specific outcome (dependent variable) that can be observed/measured in response to the experimental conditions applied as part of the treatment (Holmes, Moody and Dine, 2011).	**c.** Outcome variables proposed are irrelevant for the proposed experimental context provided or with the hypothesis. **d.** Stated outcome not measurable. **e.** No measure was proposed for the outcome variable. **f.** An outcome variable was not listed for an investigation.
		g. There is a mismatch between what the investigation claims to test and the outcome variable.
(4) Accounting for Variability	Experimental design needs to account for the variability occurring in the natural biological world. Reducing variability is essential to reduce effect of non-relevant factors in order to carefully observe effects of relevant ones (Box *et al.* 2005; Cox and Reid 2000).	**a.** Claims that a sample of experimental subjects will eliminate natural variability with those subjects.
	Selection of a random (representative) sample: A representative sample is one where all experimental subjects from a target demographic have an equal chance of being selected in the control or treatment group. An appropriate representative sample size is one that averages out any variations not controlled for in the experimental design. (The College Board, 2006; Holmes, Moody and Dine, 2011).	**b.** Criteria for *selecting* experimental subjects for treatment vs. control group are biased and not uniform. **c.** Criteria for selecting experimental subjects for investigation are different in a way that is not representative of the target population.

Table 2.2: Rubric for Experimental Design (RED).

Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
	Randomized design of an experiment: Randomizing the order in which experimental subjects or units experience treatment conditions as a way to reduce the chance of bias in the experiment (Ramsey and Schafer, 2012). Randomization can be complete or restricted. One can restrict randomization by using block design which accounts for known variability in the experiment that can't be controlled.	**d.** Decisions to *assign* experimental subjects to treatment vs. control group are not random but biased for each group.
		e. Random assignment of treatments is not considered. **f.** Random assignment of treatments is incomplete as they show random assignment of the experimental subjects but instead, what is needed is random assignment of treatments.
	Replication of treatments to experimental units or subjects: Replication is performed to assess natural variability, by repeating the same manipulations to several experimental subjects (or units carrying multiple subjects), as appropriate under the same treatment conditions (Quinn and Keough, 2002).	**g.** Replication means repeating the entire experiment *at some other time* with another group of experimental subjects. **h.** No evidence of replication or suggested need to replicate as a method to access variability or to increase validity/power of an investigation.
(5) Scope of Inference of Findings	**Scope of inference:** Recognizing the limit of inferences that can be made from a small characteristic sample of experimental subjects or units, to a wider target population and knowing to what extent findings at the experimental subject level can be generalized.	**a.** The inference from a sample is to a different target population. Usually students under- or overestimate their findings beyond the scope of the target population. **b.** No steps are carried out to randomly select experimental subjects' representative of the target population about which claims are made.
	Cause and effect conclusions: A cause-and-effect relationship can be established as separate from a mere association between variables only when the effect of lurking variables are reduced by random assignment of treatments and matching treatment and control group conditions as closely as possible. Appropriate control groups also in comparison to the treatment group also need to be considered (NIST/SEMATECH, 2003; Wuensch, 2001).	**c.** A causal relationship is claimed even though the data shows only association between variables. Correlation does not establish causation. (NIST/SEMATECH, 2003)

Refer to Appendix F for 'Glossary of Terms'

2.5.2 Difficulties with Experimental Design Detected Using the Published Assessments (RQ2)

To understand, *to what extent published assessments reveal evidence of first-year undergraduate biology students' knowledge and difficulties with experimental design,* we used responses to the 'Shrimp', 'Drug', and 'Bird' assessments to identify students' correct ideas and difficulties which, as shown in Table 2.3, were then classified within all 5 categories of difficulty. In the following sections, we discuss the examples of student responses from Table 2.3, demonstrating correct ideas and typical difficulties with five RED areas to each of three assessments. Detailed explanations of each example are provided. For each assessment, a more complete example from a student with an overall correct idea and a typical response from a student that shows difficulties are presented in supplementary information Tables SI 1-3. For confidentiality, pseudonyms are used to identify each student.

2.5.2.1 Variable Property of an Experimental Subject

Difficulty with identifying an appropriate experimental subject with a variable property to be investigated was a problem for students across all three assessments. Students had trouble recognizing that an experimental subject possesses properties that vary, the sample of experimental subjects must display an appropriate variable property aligned with the given experimental goal, and the variable property needs to be consistently considered when planning an investigation (Table 2.2; Area of Difficulty 1 a-c).

As illustrated in Table 2.3 (1.Shrimp.C), Anna correctly recognizes tiger shrimp as an experimental subject in the 'Shrimp Assessment', but Beth shows a difficulty with the experimental subject (tiger shrimp) as she considers it to be a variable and includes it as a part of the experiment control (1.Shrimp.D). Instead, the correct idea would be to think of a *variable property of the experimental subject* (Table 2.2; Area of Difficulty1a).

In the 'Drug Assessment', Josh suggests maintaining the variable property "blood pressure" constant (Table 2.3, 1.Drug.C) but Ken proposes experimental subjects

divergent from the proposed target population (Table 2.2, Area of Difficulty 1-b). This is a problem because Ken considers including patients on the basis of pregnancy status and age (1.Drug.D) instead of sampling an appropriate target population for the drug (people with high blood pressure).

For the 'Bird Assessment', one appropriate variable property of birds is the species: blackbirds, zebra finches and geese. Part of the assessment asks about differences in food preference for zebra finches but another part focuses on one gender (male) of three different bird species. Rita considers the experimental subject (birds) appropriately with reference to the gender of zebra finches in her initial response and then she proposes a study with the three species but maintains a consistent reference to the birds' gender (Table 2.3, 1.Bird.C.). This shows that Rita correctly explains the *experimental subject* in terms of a *variable property* aligned with the goal of the experiment. In contrast, Sara, in the first part of the response, considers groups of experimental subject based on the gender of zebra finches. But then she shifts to talking about the species with no reference to a specific gender (1.Bird.D.). This shows a lack of coherence because variable property of the experimental subject was not consistently considered (Table 2.2, Area of Difficulty 1-c).

Table 2.3: Examples of student responses with the RED areas of difficulty across three assessments.

---

1. Variable property of an experimental subject
'Shrimp Assessment'
Correct (C) idea from Anna: *"The advantage to having only tiger shrimp in the experiment is that you are only using one single species of shrimp. This leads to an advantage because there is less variability within the growth of shrimp."*
Difficulty (D) from Beth: *"The tiger shrimps act as the control group."* (Area of Difficulty 1-a)
'Drug Assessment'
Correct (C) idea from Josh: *"Patients need to have [same range of] high blood pressure."*
Difficulty (D) from Ken: *"Participants cannot be pregnant simply because it will affect the fetus differently than the adult. People older than 35 should not test the drug…"* (Area of Difficulty 1-b)
'Bird Assessment'
Correct (C) idea from Rita: *"…Knowing from previous research that male birds do not avoid solid colors…" […] Ensuring that all of the birds being tested are as similar as possible except for the treatment is best. This entails that all birds have the same gender…"*
Difficulty (D) from Sara: *"The reason for these differences between the two sexes could have to do with the fact that one sex is the main contributor of food to their young." […] You could set up three separate areas having one species assigned to one of the three."* (Area of Difficulty 1-c)

---

2. Manipulation of variables
'Shrimp Assessment'
Correct (C) idea from Anna: *"1. A Low salinity; 2. A high salinity; 3. B low salinity; 4. B high salinity; 5. C low salinity; 6. C high salinity."*
Difficulty (D) from Beth: *"…Low salinity with no nutrient, high salinity with no nutrients…"* (Area of Difficulty 2-c; 2-f)
'Drug Assessment'
Correct (C) idea from Josh: "[Administration of] new drug… […] *lower the blood pressure of people with high blood pressure to a safe level."*
"…same range of high blood pressure, diet, exercise, eating habits, sleep habits…"
Difficulty (D) idea from Ken: (i) *"This drug will be administered to people at low dosages at first, then we will record results and from there calculate the correct amount of Alamain that should be given to each person."* (Area of Difficulty 2-b)
(ii)*"Experimental groups will receive a couple of different dosages to see how each dose affects blood pressure"* (Area of Difficulty 2-d)
(iii) *"The younger, healthier participants will be the experimental group while the not so young will be the control."* (Area of Difficulty 2-j)
'Bird Assessment'
Correct (C) idea from Rita: (i) *"…each species of bird would be randomly divided into two groups, with one group receiving treatment 1 and the other group receiving treatment 2 (that is, 50 blackbirds would receive treatment 1, 50 blackbirds would receive treatment 2, and likewise for zebra finches and geese)…."*
(ii) *"Ensuring that all of the birds being tested are as similar as possible except for the treatment is best. This entails that all birds have the same gender, are roughly the same age, come from very similar habitats, and are in overall good health (no underlying conditions such as currently suffering from a given disease)."*
Difficulty (D) idea from Sara: (i) *"You could repeat the experiment but this time allowing all three of the species to be in the same area."* (Area of Difficulty 2-d; 2-f)
(ii) *"…this experiment would take into account any competition [among all three bird species] that might take place"* (Area of Difficulty 2-g)

---

3. Measurement of outcome
'Shrimp Assessment'
Correct (C) idea from Anna: *"…The advantage to having only tiger shrimp in the experiment is that there is less variability within the growth of a single species of shrimp."*

Table 2.3: Examples of student responses with the RED areas of difficulty across three assessments.

Difficulty (D) from Beth: *"a researcher can confidently expect to find a repetitive response to a given exposure in a group of genetically identical tiger shrimps..."* (Area of Difficulty 3-e)

'Drug Assessment'

Correct (C) idea from Josh: *"If people who take the drug consistently have decreased blood pressure, then the drug is effective."*

Difficulty (D) from Ken: *"If the drug does indeed reduce blood pressure, the percentage of those whose blood pressure [becomes] normal will be significantly higher than that control group."* (Area of Difficulty 3-g)

'Bird Assessment'

Correct (C) idea from Rita: *"...differences in the response variable (in this case, the frequency of avoiding or not avoiding food given the particular treatment) can be [attributed to] the difference in treatment."*

Difficulty (D) from Sara: *"...they [all three bird species] all will be in the same area together and not separated.... This would increase the power by determining which seed the birds compete over and which seed the birds ignore [...] After the time is up, you could collect the remaining seeds and see which treatment was eaten the most and which treatment the birds avoided the most."* (Area of Difficulty 3-c; 3-g)

---

4. Accounting for variability

'Shrimp Assessment'

Correct (C) idea from Anna: "...using only tiger shrimps reduces variance..."

"...there are two tanks with each treatment..."

"In order for randomization to occur it might be easiest to use dice and assign each number to its corresponding treatment number. Example: Roll dice 1+ 2; Outcome Die 1= 2 and Die 2= 4. From this you would put treatment two and four in tanks 1 and 2."

Difficulty (D) from Beth: (i) *"...a researcher can confidently expect to find a repetitive response to a given exposure in a group of genetically identical tiger shrimps."* (Area of Difficulty 4-a; 4-h)

(ii) *"With all the shrimp in one tank, one by one randomly assign a shrimp to a tank [...] by doing this, the biologist is aware of which tanks contain which ingredients but the shrimp are completely randomized."* (Area of Difficulty 4-f)

'Drug Assessment'

Correct (C) idea from Josh: *"They [experimental subject/participants] will have to be at the same range of high blood pressure, diet, exercise, eating habits, sleep habits."*

"They [participants] will be chosen at random to be part of the experimental or control group that way they do not have an opinion on how the drug may or may not be helping them."

Difficulty (D) idea from Ken: (i) *"People older than 35 should not test the drug. These criteria need to be met and not taken lightly because health problems may arise."* (Area of Difficulty 4-c)

(ii) *"The younger, healthier participants will be the experimental group while the not so young will be the control."* (Area of Difficulty 4-d)

'Bird Assessment'

Correct (C) idea from Rita: *"...each species of bird would be randomly divided into two groups, with one group receiving treatment 1 and the other group receiving treatment 2...."*

Difficulty (D) from Sara: *"You could set up three separate areas having one species assigned to one of the three."*

(Area of Difficulty 4-e)

---

5. Scope of inference

'Shrimp Assessment'

Correct (C) idea from Anna: *"One statistical disadvantage to only having only tiger shrimp is that due to the fact we only used one species of shrimp we are not able to make a generalization about all shrimp."*

Difficulty (D) from Beth: *"...this fails to demonstrate how a given ingredient may affect another type of shrimp. Ultimately it limits the depth of the study."* (Area of Difficulty 5-b; 5-c)

'Drug Assessment'

Correct (C) idea from Josh: *"participants with same range of high blood pressure, diet, exercise, eating*

Table 2.3: Examples of student responses with the RED areas of difficulty across three assessments.

habits, and sleep habits." "...blood pressure [will be measured]." "...participants chosen at random..."

Difficulty (D) from Ken: *"...health, hemoglobin, smoking, age under 35, and pregnancy status..."* (Area of Difficulty 5-a; 5-c).

'Bird Assessment'

Correct (C) idea from Rita: *"...With all of these potential differences eliminated, the birds would be made different in only one respect: their treatment. In this manner, one would be able to confidently declare that differences in the response variable [in this case, the frequency of avoiding or not avoiding food given the particular treatment] can be laid at the feet of the difference in treatment."*

Difficulty (D) from Sara: *"The reason for these differences between the two sexes could have to do with the fact that one sex is the main contributor of food to their young." [...] You could set up three separate areas having one species assigned to one of the three." "...determining which seed the birds compete over and which seed the birds ignore" "You could set up three separate areas having one species assigned to one of the three."* (Area of Difficulty 5-b; 5-c).

2.5.2.2 <u>Manipulation of Variables</u>

Across the three assessments, an appropriate response for manipulating variables would have been to come up with appropriate treatment and control groups and to recognize unrelated variables to a given study. A clear pattern of difficulties was found across the three assessment instruments when students were challenged to hypothesize and manipulate treatment variables during the process of experimental design. Students often did not focus on the right variables. Sometimes they considered irrelevant variables while other times they proposed inappropriate treatments or failed to combine two treatments as required for the experimental goal. Finally, students had trouble matching treatment and control conditions to neutralize effects of lurking/confounding variables for an experiment (Table 2.2; Area of Difficulty 2 a-j).

With the 'Shrimp Assessment,' Anna sets up appropriate treatment groups carrying combinations of two independent treatment variables (nutrient and salinity) applied to the experimental subject (tiger shrimp) (Table 2.3, 2.Shrimp.C.). However this seems to be difficult for Beth who haphazardly proposes treatment groups (Table 2.2, Area of Difficulty 2-c) with missing conditions to keep the shrimp alive (2.Shrimp.D.). This also shows a problem with *combinatorial reasoning* as Beth fails to combine salt and nutrients appropriately to find their effect on the growth of shrimp (Area of Difficulty 2-f).

Josh's hypothesis for the 'Drug Assessment' shows a clearly predicted testable association between a treatment and outcome (Table 2.3, 2.Drug.C.). In contrast, Ken demonstrates a difficulty in framing a hypothesis as he fails to identify a clear expected result from the proposed experiment, as evident from 2.Drug.Di (Table 2.2, Area of Difficulty 2-b). Also, Ken proposes treatment conditions like "different dosages of the blood pressure drug" (2.Drug.Dii.) inappropriate to the original goal of the investigation, which is to test effect on blood pressure from the presence and absence of drug intake (Table 2.2, Area of Difficulty 2-d). In an experiment, the control and experimental groups are required to be matched as closely as possible to equally reduce the effect of unrelated variables on both groups. Josh demonstrates this ability well by matching appropriate

variables to control lurking variables in a study to develop a high blood pressure drug (2.Drug.C.). However, Ken should not have assigned the participants (experimental subjects) carrying obvious differences (young/healthy and not so young) to treatment and control group, respectively (2.Drug.Diii.) (Table 2.2, Area of Difficulty 2-j), because parameters other than the treatment variables need to be identical for both the treatment and control conditions.

For the 'Bird Assessment,' Rita correctly organizes assignment of experimental units to treatments in alignment with the experimental goal to examine preference in consuming either of two kinds of pesticide granules among three different bird species separated by a block design (Table 2.3, 2.Bird.C.). Sara on the other hand, tries to combine all three different bird species within a single treatment group (2.Bird.Di.) when instead, the effect of treatments are to be determined individually for each bird species by "block design." Thus we conclude Sara shows a difficulty in identification of treatment groups and combinatorial reasoning (Table 2.2, Area of Difficulty 2-d; 2-f).

Another measure to identify treatment and control groups by Rita was controlling outside variables by matching up the various treatment groups in terms of lurking variables that could affect bird behavior (Table 2.3, 2.Bird.C.). In contrast, Sara considers "competition among bird species" as a variable which is unrelated to the intended goal of finding out what pattern or color of pesticide granules each species would avoid eating (2.Bird.Dii.) (Table 2.2, Area of Difficulty 2-g).

2.5.2.3 Measurement of Outcome

With correct knowledge of measurement *of outcome*, a student would propose experimental outcomes using appropriate measures. However, in their responses to all three assessments, some students struggled with measures when they either failed to state outcomes that were measurable or they proposed outcomes without specific measures in terms of units or categories. Sometimes those that did propose measurable outcomes suggested variables that were mismatched to a given experimental goal (Table 2.2; Area of Difficulty 3 a-g).

The "growth of shrimp" as a measurable outcome is correctly identified in Anna's response to the 'Shrimp Assessment' (Table 2.3, 3.Shrimp.C.) But for Beth's response (3.Shrimp.D.), the phrase "repetitive response" provides no measure for a specific outcome thereby she demonstrates difficulty for measurement of outcome (Table 2.2, Area of Difficulty 3-e).

For the 'Drug Assessment', Josh suitably suggests "decrease in blood pressure" as outcome (Table 2.3, 3.Drug.C.). But Ken's proposed outcome (3.Drug.D.) illustrates a mismatch between the goal of the investigation and the outcome to be measured (Table 2.2, Area of Difficulty 3-g). Specifically, this is a mismatch because having more participants with normal blood pressure is different from saying that participants' blood pressure will be lower if the drug is effective.   In other words, an effective drug is one that simply reduces high blood pressure for the treatment group participants but not necessarily down to normal levels.

In the 'Bird Assessment', an appropriate measure for an outcome variable is suggested by Rita (Table 2.3, 3.Bird.C.). Sara shows a problem with her proposed *measurement of outcome* (3.Bird.D.) when she indicates that the bird species will "compete" for seeds, which is irrelevant to the stated goal of this investigation (Table 2.2, Area of Difficulty 3-c). There is a mismatch between what the question asked and the investigation goal because "which treatment was eaten the most" is not a relevant outcome when the goal is to find out whether or not the birds consume seeds, not "how much" they consume (Area of Difficulty 3-g).

2.5.2.4 Accounting for Variability

Correct ideas about accounting for variability would require recognizing natural variation among experimental subjects while trying to reduce variation sourced externally from unrelated factors. We found that across three assessments students showed flawed ideas concerning variability in multiple ways. Either they completely failed to recognize natural variation or they failed to account for variability with appropriate methods like replicating and randomizing treatment assignments (Table 2.2; Area of Difficulty 4 a-h).

For the 'Shrimp Assessment', Anna shows a correct understanding of how to deal with natural biological variability (Table 2.3, 4.Shrimp.C.). In contrast, Beth reveals a difficulty with variability (4.Shrimp.Di.) as the phrase *"genetically identical tiger shrimps"* incorrectly claims that having only tiger shrimp eliminates natural variability. In fact, some variability exists even within a sample of the same species (Table 2.2, Area of Difficulty 4-a). Another component for this area includes *replication of treatment conditions* as a measure to assess natural variability within an experimental unit carrying multiple experimental subjects. This is included in Anna's response (4.Shrimp.C.), but Beth does not consider replication of treatment (4.Shrimp.Dii.) (Table 2.2, Area of Difficulty 4-h).

To account for known variability from lurking variables in an experiment requires randomizing the order in which experimental units experience treatment conditions (Table 2.2, Area of Difficulty 4). Randomization is well described in Anna's response as she illustrates a complete randomization of assignment of both treatment and shrimps to tanks (Table 2.3, 4.Shrimp.C.). Alternatively, an incomplete randomization procedure (Table 2.2, Area of Difficulty 4-f) is suggested by Beth who only randomizes assignment of shrimp to tanks but fails to randomize assignment of treatment combinations to each tank (Table 2.3, 4.Shrimp.Dii.).

For the 'Drug Assessment', Josh proposes to deal with variation using a random sample to represent a target population (Table 2.3, 4.Drug.C.). Instead, Ken selects experimental subjects that are not representative of the target demographic population and are also not randomly chosen (Table 2.2, Area of Difficulty 4-c) (4.Drug.Di and ii.), because participants with different characteristics are purposefully assigned to treatment and control groups (Table 2.2, Area of Difficulty 4-d).

In the 'Bird Assessment', evaluating how students randomly assign each of three bird species to two treatments provides a measure of how well students address natural variability in an experiment. This is demonstrated well by Rita (Table 2.3, 4.Bird.C.). Alternatively, Sara sets up separate areas for each species but does not specify how

treatments are assigned in a randomized fashion (4.Bird.D.) (Table 2.2, Area of Difficulty 4-e).

2.5.2.5 Scope of Inference

When a student demonstrates correct ideas about interpretation of experimental findings they estimate an appropriate extent of inference of findings and are also able to draw logical causal claims. But across the three assessments, we found students went wrong with interpretation of experimental findings in several ways. They either over, or under-estimated experimental claims, or they made inappropriate inferences about causal relationships while their experimental procedures only suggested correlation among variables (Table 2.2; Area of Difficulty 5 a-c).

For the 'Shrimp Assessment', both Anna and Beth recognize the limit of inferences from a small sample of tiger shrimps (Table 2.3, 5.Shrimp.C.). However, Beth still shows difficulty in this area because she does not mention a measurable outcome or randomization and replication of treatments and fails to recognize natural variability with the experimental subjects. With such flaws, Beth only show signs of correlation and not causal association (5.Shrimp.D) between application of variable nutrient and salinity conditions and growth of tiger shrimps (Table 2.2, Area of Difficulty 5-b; 5-c).

On the 'Drug Assessment', Josh's experimental findings can be generalized to an appropriate sample of the target population of people with high blood pressure. He makes specific considerations during selection of experimental subjects and the identification of experimental groups, and he applies methods to deal with variability (Table 2.3, 5.Drug.C.). Similarly, his proposed *measurement of outcome* ("blood pressure") and measures for *accounting for variability* ("participants chosen at random") justify appropriate cause and effect conclusions about the effectiveness of the high blood pressure drug. In contrast, Ken's study will apply to a different target population and not the intended subjects with high blood pressure due to lack of appropriate *accounting for variability* measures and a skewed participant pool with demographic properties that are not representative of a larger target population (Table 2.2, Area of Difficulty 5-a).

Similarly, due to selection bias based on irrelevant variables (5.Drug.D.), when he selects and assigns participants to treatment groups, causal claims would be inappropriate because of Ken's flawed comparison groups (Area of Difficulty 5-c).

For the 'Bird Assessment', careful considerations include appropriate groups of experimental subjects, an organized set up of experimental groups, suitable measurable outcomes, and methods to account for natural variability among bird species for Rita's study, making her design suitable for causal claims. Rita correctly asserts a causal claim in her answer (Table 2.3, 5.Bird.C.). In contrast, Sara's experimental design lacks coherence in several areas. The experimental groups are not considered consistently across different parts of the response, treatment assignments follow a pattern unsuitable to the study goal, proposed outcomes do not match the original investigation goal, and efforts to account for natural variability are inadequate. These flaws make it unfeasible to draw any *cause and effect conclusions* (5.Bird.D.) from Sara's experimental proposal (Table 2.2, Area of Difficulty 5-b; 5-c).

2.5.2.6 Interconnectedness of RED Areas of Difficulty

In examining problems with student interpretation of experimental findings for each of the three assessments, an interesting finding was that student difficulties with two RED categories (Tables 2 and 3) often went together. The categories were not independent but interconnected. For example, it is not surprising that a difficulty with controlling outside variables categorized under *manipulation of variables* was associated with difficulty *accounting for variability* because controlling outside variables provides a way to account for and minimize natural variation in samples. Likewise, proposal of a suitable testable hypothesis with appropriate *manipulation of variables* was connected to *measurement of outcome* difficulties because if the hypothesis carried inappropriate relationships between treatment and outcome variables, the outcome measurements were also flawed. *Accounting for variability* influenced inferences drawn from experimental findings or *scope of inference*. Without considering variability, students overestimated or underestimated findings beyond the scope of the participating sample of a "population" in a study (Table 2.2, Area of Difficulty 4-a). Similarly, correlations were erroneously

considered to demonstrate experimental evidence for causal relationships. Causation requires possible lurking variables to be carefully controlled for by random selection of representative experimental subjects.

The various types of "Typical Evidence of Difficulties" in the RED (Table 2.2) were confirmed with responses to three different assessments as illustrated with quotes (Table 2.3). Supplementary Information (SI) Tables SI 1-3 provide actual student responses with examples of typical correct ideas and difficulties according to the RED. The difficulties are underlined and coded with a footnote that corresponds to Table 2.2. But the examples discussed did not illustrate all types of "Typical Evidence of Difficulties" from Table 2.2, so actual responses to illustrate other difficulties are provided in Tables SI 4-6. Consistently, a careful analysis of responses revealed difficulties with experimental design in five areas: *(1) a property of an experimental subject that is variable; (2) manipulation of variables; (3) measurement of outcome; (4) accounting for and measuring* variability and (5) *scope of inference of findings*. These five areas were used to develop the RED and thus formed the foundation for subsequent analysis.

2.5.3 Efficacy of the RED to Detect Changes in Students' Experimental Design Abilities (RQ3)

With the various experimental design difficulties now characterized in the RED, we recognized that for practical purposes, RED must be validated for its usefulness to detect changes in undergraduate student responses before and after a course (RQ3). We argued that, if RED is sensitive enough to detect changes in the proportion of undergraduate students with correct responses, a similar measure at the end of course would help us find out if students are learning about experimental design from our course. To make good decisions about how to focus on student difficulties that needed attention, we needed to know if some assessments were better than others at probing particular knowledge. The proportion of students that showed correct ideas or difficulties was calculated after coding responses with the RED. For each area, the percentage of students with correct knowledge (dark gray), difficulties (medium gray), or lack of

evidence (light gray) is presented in Figure 2.2. Results show that with the three selected assessments, RED coding is capable of detecting differences in the proportion of students with correct knowledge or difficulties in the five experimental design areas (Table 2.2).

Our analysis showed that in case of certain RED areas, there were significant differences between pre- and posttest with p-values ranging from ≤0.01 to ≤0.1, which implies that each assessment was capable of measuring changes in student knowledge with respect to certain RED areas. We consider p<0.1 significance level to be adequate because with written response data, our understanding of changing knowledge is limited to what students write. Thus, we might have a 10% chance of being uncertain about the precision of these assessments in demonstrating experimental design knowledge. However, for research purposes with a cut off at p<0.05 significance levels, each assessment would still be a useful measure of certain RED areas. For example, the 'Shrimp' and 'Drug Assessment' reports pre vs. post p-values for areas like *variable property of experimental subject* at <0.05 significance levels.
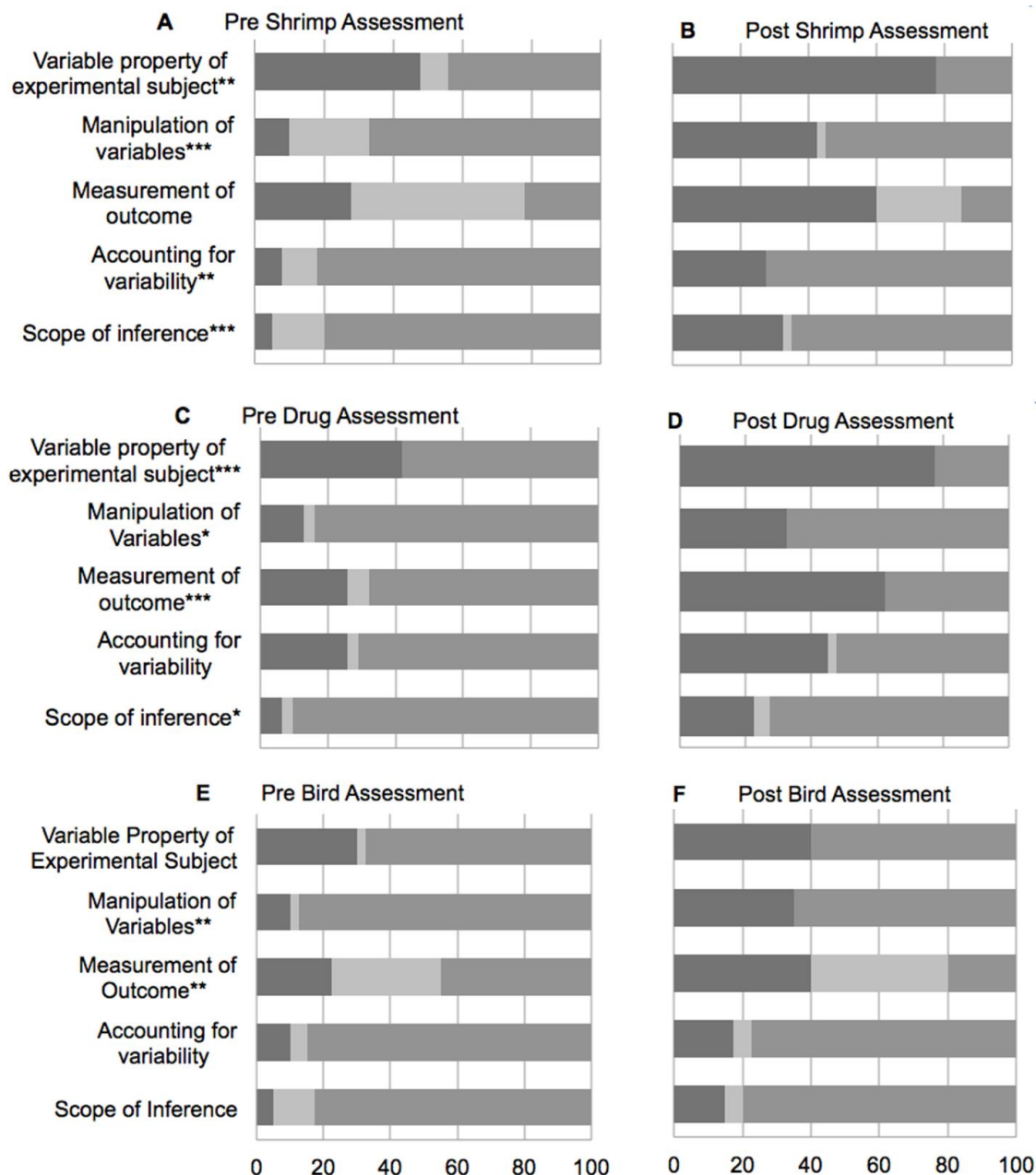
Figure 2.2: Proportions of Students with Correct Ideas, Difficulties and Lack of Evidence for Knowledge of Experimental Abilities

Proportions of students who had correct ideas (dark gray), difficulties (medium gray) and lack of evidence (light gray) for knowledge of experimental abilities as probed by three assessments administered at the beginning and at the end of a semester. The 'Shrimp Assessment' was given as a post-test during 2009 to cohort A (Panel B; n= 40) and as pre-test in the following year during 2010 to cohort B (Panel A; n=40). The 'Drug Assessment' was used as a post-test in 2011 to cohort C (Panel D; n=40) and as a pre-test in 2012 to cohort D (Panel C; n=31). The 'Bird Assessment' was assigned as post-test in 2010 to cohort B (Panel F; n=40) and as a pre-test in 2011 to cohort C (Panel E; n=40). The y-axis topics are 'Areas of Difficulty' from Table 2.2.

Looking across the data for the three assessment instruments (Figure 2.2), a clear pattern of differences at the beginning and end of a course is revealed when RED was used to code a sample of responses. The *manipulation of variables* is an area that consistently showed significant difference between the pre- and post-test for all three assessments. This difference was detected even though, for all three assessments, more than half of the students still showed difficulty with *manipulation of variables* at the end of the course. Figure 2.2 shows that even though a significant difference was not found on one of the tests for *variable property of an experimental subject*, *measurement of outcome*, and *scope of inference*, the trend was the same as for two of the assessments that did show a significant difference at the beginning and end of a course in these areas. Although one area showed significant difference between the pre- and post-test for only one assessment, *accounting for variability* trends were also similar for this area across all three tests.

All three assessments showed similar differences in the proportion of students with correct ideas about experimental design and the areas of difficulties that need to be addressed. Next we present Figure 2.2 findings, first in terms of the magnitude and direction of change in the proportion of students with correct ideas about experimental design, and then by considering the proportion of students who have difficulties in each area when responses are coded using the RED.

The proportion of students with correct responses at the beginning and the end of the course are aligned for all areas across three assessments in Figure 2.2 A-F. For the 'Shrimp Assessment', by the end of semester *variable property of experimental subject, manipulation of variables*, and *measurement of outcome* showed the largest differences in proportion of students with 'correct' ideas (Figure 2.2 A-B) (Supplementary Information Table SI 11 shows actual differences in proportion of students with ideas that were 'correct' or showed 'difficulty' at the beginning or end of a semester with each assessment). Similarly, the 'Drug Assessment' showed more differences in 'correct' responses for *variable property of experimental subject* and *measurement of outcome,* but it was less sensitive for detecting differences in the proportion of students with correct

ideas for *manipulation of variables* (Figure 2.2 C-D). The 'Bird Assessment' was most sensitive in detecting pre to post differences in the proportion of students with 'correct' ideas in the areas, *manipulation of variables* and *measurement of outcome,* but it was less sensitive for prompting correct ideas about *variable property of experimental subject* at the end of the course (Figure 2.2 E-F). A small portion of students had correct ideas about *accounting for variability* at the end of the course except for with the 'Drug Assessment', which similarly prompted nearly a fourth of the students to account for variability at the start of the course. Differences were small but the trend was the same across all three assessments. According to all three assessments, although some differences are apparent, only a small portion of students had correct ideas about *scope of inference* even at the end of the course. We would like to acknowledge that since the assessments were used for diagnostic purposes, we did not give partial credit for distinguishing average students from those with poor understanding, corresponding to each RED area. A relatively stringent cut off was appropriate because we did not use their responses to grade students. The assessments simply provided opportunities for them to demonstrate their thinking so we would know what the problems are when students design experiments.

In addition to detecting correct ideas, each assessment also captured information about the proportion of students who demonstrated 'difficulties' with five experimental knowledge areas. From the beginning to the end of the semester, the 'Shrimp Assessment' measured the largest differences in 'difficulty' for *variable property of experimental subject* and *scope of inference* but for *measurement of outcome* the difference found was only 8% (medium gray bars in Figure 2.2 A-B). For the 'Drug Assessment', the biggest differences in proportion of students with 'difficulty' were detected for *variable property of experimental subject* and *measurement of outcome* and it was less sensitive for detecting difference in difficulties for *manipulation of variables* (medium gray bars in Figure 2.2 C-D). Similarly, for the 'Bird Assessment', the largest differences in the proportion of students with 'difficulties' were found for the areas, *measurement of outcome* and *manipulation of variables,* while difficulties involving *accounting for variability* and *scope of inference* remained almost unchanged at the end of semester (medium gray bars in Figure 2.2 E-F). Note that all three assessments were

good at exposing students' difficulties in the five areas, which is useful for students and the instructor to know so that the problems can be fixed.

An assessment with a large portion of 'lack of evidence' responses is less useful for diagnostic purposes. The 'Drug Assessment' showed the lowest prevalence of lack of evidence responses (light gray bars in Figure 2.2C-D). The *measurement of outcome* area was most problematic for 'lack of evidence' on both the 'Shrimp Assessment' and the 'Bird Assessment' (light gray bars in Figure 2.2A-B and 1 E-F).

In general, looking across the three assessments the areas, *variable property of an experimental subject* and *measurement of outcome,* were easier for most students at the end of the course than *manipulation of variables*, *accounting for variability* or *scope of inference.* However, *variable property of an experimental subject* for the 'Bird Assessment' was harder than for the 'Shrimp' and 'Drug Assessment'. Also, the 'Bird Assessment' did not probe well for *measurement of outcome*. *Accounting for variability* was slightly easier in the 'Drug Assessment' than the 'Shrimp' and 'Bird Assessment' perhaps because the 'Drug Assessment' specifically probes for ways to deal with variability like selecting a representative sample and randomized design of an experiment (Table 2.2; Area of Difficulty 4). A reason why *accounting for variability* was more difficult with the other assessments could be that the assessments did not guide students to address variability. Finally, it is interesting to note that *scope of inference* was problematic for students according to all three assessments even though a slightly larger proportion of students demonstrated correct ideas in this area at the end of the course for all three assessments (Figure 2.2A-F; Row 5).

2.6 Discussion

In summary, our study yielded the following major findings:

1. All *Established* difficulties documented in our literature review (Table 2.1) were consistently found in responses from our own undergraduate biology students;

2. Data from our undergraduate biology students permitted the re-classification of one *Partially Established* difficulty, the *variable property of experimental subject*, to *Established*;

3. Data collected from undergraduate biology students, together with difficulties data from a review of the literature, confirmed five major areas of difficulty with experimental design: *(1) a property of an experimental subject that is variable; (2) manipulation of variables; (3) measurement of outcome; (4) accounting for and measuring variability and (5) scope of inference of findings*;

4. All the above data was used to inform the development of a Rubric for Experimental Design (RED), consisting of descriptions of correct ideas and typical difficulties within each of the abovementioned 5 major areas;

5. The RED was shown to be an effective tool for detecting changes in undergraduate students' experimental design knowledge during instruction.

In response to RQ1, our comprehensive literature review (Table 2.1) summarized for the first time the full range of published experimental design difficulties and classified 5 categories and 13 sub-categories of difficulty on a framework that told us whether they required further research or not in order to be fully identified. In fact, nearly all reported difficulties were confirmed to be fully *Established* and therefore ready to be incorporated into our rubric. The one *Partially Established* difficulty, to do with *variable property of experimental subjects*, had previously been identified in only one study by Salangam (2007) with undergraduate biology students who were not science majors. We then re-classified this difficulty as *Established* from data obtained when addressing RQ2 and thus we had a full complement of all the known difficulties for our rubric.

In addressing RQ2, our undergraduate biology students demonstrated the full range of difficulties documented in Table 2.1, confirming the important need to address such difficulties in instruction. Indeed we were concerned to find that several of the experimental design difficulties identified as long as 50 years ago by Karplus (Fuller,

2002) still persist today among our students. In addition, a difficulty with *scope of inference*, previously reported by Chen and Klahr (1999) in a study involving elementary school level students was shown by us to persist as a problem at undergraduate level. All the above findings convinced us of the important need to develop the RED that could serve as an important tool for assessing students in this crucial area of biological expertise while also informing intervention and remediation strategies.

To answer RQ3, RED was then used in a pre-/post-test comparison of experimental design knowledge and difficulties to find out if it can be usefully deployed with published assessments to discriminate changes in knowledge during course participation. RED was found to be useful with all three assessments. The RED further helped us organize the changes in student knowledge according to five areas of difficulty. The scoring process we employed to discriminate changes before and after the course can be applied for practical purposes. Although we gathered hundreds of responses at the beginning and end of each semester from four cohorts, our random sample of 40 responses was sufficient to successfully demonstrate changes in students' knowledge. During scoring, for research purposes, we scored students for evidence of difficulties in an all or none manner. However these assessments were low stakes and provided students a forum to express their ideas freely. Alternatively, an instructor might decide to assign partial credit to let students know where they stand on a continuum.

Once developed, the RED made it possible to evaluate the strengths and weaknesses of the three assessment instruments (Figure 2.2). For example, we now know that the 'Bird Assessment' was more difficult for students in this study, perhaps because the context, ecological behavior, was not covered in this particular course (Clase *et al.,* 2010). Prior knowledge such as "competition among species" in this study can lead students astray. Lack of knowledge about the context may also lead to "lack of evidence" responses. An assessment with a high frequency of "lack of evidence" responses could potentially be improved by providing background information so that all students designing an experiment start with the same contextual knowledge. We do not know whether students who show 'lack of evidence' with *manipulation of variables* in fact had

difficulties and thus chose to not write much. Other areas with 'lack of evidence' problems on the pre-test showed a decline in 'lack of evidence' for the post test, indicating that the problem may reflect how much students chose to write in their response rather than indicating a flawed probing design for the assessment. By more specifically probing for the lack of evidence, as directed by RED, students would be better prompted to reveal their knowledge. In contrast, the other two assessment instruments performed better than the 'Bird' instrument for the sample of biology students tested in this study. Now that we can use RED to consistently grade student knowledge and to help them recognize and address their difficulties, it will be useful to gather a collection of assessments that specifically address each aspect of RED.

An alternative explanation for why students struggle with identifying components of experimental design in an unfamiliar context could be that novice students, unlike experts, frequently have trouble identifying two problems as having the same theoretical features if the context is changed (Chi, Glaser, and Farr, 1988). It is especially important to determine if students are having trouble because they lack knowledge about experimental design concepts as defined in our glossary (See Supplementary Information) or if they know about experiments, but have trouble applying what they know in an unfamiliar context. In other words, certain features might allow students to call on particular knowledge about experiments in one domain, but they may have trouble transferring what they know to a completely different domain (Barnett and Ceci, 2002; Chen and Klahr, 1999). To resolve this uncertainty, more research is needed with additional experimental design assessments.

We envision the RED being potentially useful, with a variety of existing assessment instruments including the three used in the present study, for measuring progress from experiential learning with laboratory courses, research internships, or Course-based Undergraduate Research Experiences (CUREs) and not just with lecture courses like in the current study. According to Laursen *et al.* (2010), undergraduate research experiences are often evaluated by faculty, and some "ask students to 'demonstrate their understanding of the processes of science' by framing a research

question, developing a hypothesis, designing an experiment to test it, analyzing real data, writing a research report, and presenting their own work. These examples were sparse, and institutional evaluation efforts were often described as poorly developed or even perfunctory." (p. 176, Laursen *et al*., 2010). The RED might be a useful guide for assessing experimental design-based assignments developed by faculty mentors who also consider the various components of experimental design appropriate for their local situation. Thus, to get a complete picture of student understanding of experimentation, multiple assessments should be applied to meet the RED criteria.

In considering the advantages that RED brings to the issue of experimental design in the classroom, this rubric makes it possible to consistently diagnose and score student experimental design knowledge with different assessments. It can guide identification of student deficiencies and difficulties in certain aspects of experimental design, and these can reveal a need for new learning objectives along with activities and remediation strategies to fix such deficiencies and difficulties. The RED can also be applied towards designing instructional strategies to alert both students and instructors as to pitfalls to avoid and areas in need of instruction to promote proficiency with experimental design. With information about student difficulties, the 'propositional statements' of the RED can be further used to help target the problems with specific instruction based on practicing experimental design tasks. The RED helped us find useful information about our own students as we strive to teach students not just knowledge of the subject matter but how biology is performed as a research endeavor. Thus the RED is useful to guide all stages of learning, including objectives and instruction in addition to assessment of experimental design.

For instructors who may want to use RED, they could track their students' development of experimental design knowledge and abilities in a few different ways. Considering the RED difficulties (Table 2, column 'Typical evidence of difficulties'), an instructor could place examples for each difficulty from Table 2.3 plus examples in supplementary information (Table SI4-6) or examples from their own students, in a scoring rubric. As examples for scoring a particular assessment, a table with difficulties

from the 'Shrimp Assessment' and 'Drug Assessment' are posted online
(http://tinyurl.com/REDShrimp and http://tinyurl.com/REDDrug). Instructors might
create their own assessment, informed by the RED, and use it to examine the quality of
their instrument. The RED outlines five major areas of difficulty and if an assessment
fails to probe for a target area, the instructor could modify the directions to convert their
own assessment into a more effective probe.

For the educational researcher the RED can be used to guide and focus the design
of educational research to do with experimental design and causal explanations because it
details the components of experiments to consider. Thus it can guide the coding of expert
and novice explanations of experimental design as well as the content analysis of
textbook portrayals of experiments, and how those impact learning. For example, biology
textbooks tend to show experiments with visualizations such as graphs. The three
assessments used in the current study had no visualizations, which was a limitation. One
way for an educational researcher to understand if experts differ from students in their
knowledge about experimental design could be to have them visualize the concepts of
their experimental design with graphs. A graph might help students organize their
approach to using experimental design concepts. Drawings like graphs might represent
the five areas of experimental design difficulties from the RED in a visual form. For
instance instructors can alert their students that the experimental subject is typically
stated in the graph legend (Table 2; Area of Difficulty 1), the x-axis represents the
treatment variables (Area of Difficulty 2) and the y-axis generally shows the measurable
outcomes (Area of Difficulty 3). Students can also be alerted to graphically make
attempts to represent the variation (Area of Difficulty 4), say in the form of error bars,
and that when interpreting a graph they should consider the sample, the controls,
treatment and outcome variables, and explain the extent to which claims can be inferred
for a given experiment (Area of Difficulty 5).

With the RED to diagnose experimental design difficulties, future research can
target specific difficulties with interventions to teach beginner researchers what to do and
what not to do by using graphs or other drawings to focus their attention on each of the

five component areas in Table 2. Clearly, much work remains to be done to help biology students understand research to meet academic standards and to gain a competitive employment edge upon graduation. We suggest biologists might use RED as a framework based on empirical evidence to guide beginner researchers to develop competence in experimental design.

<u>References</u>

AAAS (2010). Vision and Change: A Call to Action, Washington, DC.

AAMC-HHMI (2009). Association of American Medical Colleges and the Howard Hughes Medical Institute, Report of Scientific Foundations for Future Physicians Committee. Washington, DC: Association of American Medical Colleges.

Association of American Colleges and Universities (2013). Scientific Thinking and Integrative Reasoning Skills. Washington, DC: Author.

Barnett SM, Ceci SJ (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological bulletin *128*, 612.

Beck CW, Blumer LW (2012). Inquiry-based ecology laboratory courses improve student confidence and scientific reasoning skills. Ecosphere *3*,112.

Box GE, Hunter JS, Hunter WG (2005). Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition. Wiley.

Bullock M, Ziegler A (1999). Scientific reasoning: Developmental and individual differences. Indiv Dev from *3*, 38-54.

Burns JC, Okey JR, Wise KC (1985). Development of an integrated process skill test: TIPS II. J Res in Sci Teach *22*, 169-177.

Chen Z, Klahr D (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. Child Dev *70*, 1098-1120.

Chi MT, Glaser RE, Farr MJ (1988). The nature of expertise. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Clase KL, Gundlach E, Pelaez NJ (2010). Calibrated Peer Review for Computer-Assisted Learning of Biological Research Competencies. Biochem Mol Biol Edu *38*, 290-295.

Cohen JA (1960). Coefficient of agreement for nominal scales. Educ Psychol Meas *20*, 37-46.

College Board (2006). AP® Statistics Free-Response Questions. http://apcentral.collegeboard.com/apc/public/repository/_ap06_frq_statistics_516 53.pdf (accessed 20 December 2013).

College Board (2009). AP® Statistics Free-Response Question Form B. http://apcentral.collegeboard.com/apc/public/repository/ap09_frq_statistics_formb .pdf (accessed 20 December 2013).

Colon-Berlingeri M, Borrowes PA (2011). Teaching biology through statistics: application of statistical methods in genetics and zoology courses. CBE Life Sci Educ *10*, 259–267.

Conference on the Psychology of Science: Implicit and explicit reasoning (2010). Purdue University. http://www.purdue.edu/hhs/psy/research/symposium/2ndPSPS2010.php (accessed December 20, 2013)

Cox DR, Reid R (2000). The Theory of the Design of Experiments. Boca Raton: Chapman and Hall/ CRC Press.

D'Costa AR, Schlueter MA (2013). Scaffolded instruction improves student understanding of the scientific method and experimental design. Am Bio Teacher *75*, 18-28.

Dolan E, Grady J (2010). Recognizing students' scientific reasoning: A tool for categorizing complexity of reasoning during teaching by inquiry. J Sci Teach Educ *21*, 31-55.

Duschl RA, Schweingruber HA, Shouse AW (2007). Taking Science to School: Learning and Teaching Science in Grades K-8. Washington, DC: The National Academies Press.

Feldon DF, Timmerman BC, Stowe KA, Showman R (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. J Res Sci Teach *47*, 1165-1185.

Fuller RG (Ed.). (2002). A love of discovery: Science education-The second career of Robert Karplus. Springer.

Gill J, Walsh J (2010). Control group. In N. Salkind (Ed.), Encyclopedia of research design. Thousand Oaks, CA: SAGE Publications, Inc

Gormally C, Brickman P, Lutz M (2012). Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments. CBE-Life Sci Educ *11*, 364-377.

Grayson DJ, Anderson TR, Crossley LG (2001). A four-level framework for identifying and classifying student conceptual and reasoning difficulties. Int J Sci Educ *23*, 611-622.

Griffith AB (2007). Semester-long engagement in science inquiry improves students' understanding of experimental design. Teach Issues Exp Ecol *5* (July 2007).

Grunwald S, Hartman A (2010). A Case-Based Approach Improves Science Students' Experimental Variable Identification Skills. J Coll Sci Teach *39*, 28-33.

Gutwill-Wise J (2001). The impact of active and context-based learning in introductory chemistry courses: An early evaluation of the modular approach. J. Chem. Ed. *78*, 684-690.

Harker AR (2009). Full application of the scientific method in an undergraduate teaching laboratory: A reality-based approach to experiential student-directed instruction. J Coll Sci Teach *29*, 97-100.

Hiebert SM (2007). Teaching simple experimental design to undergraduates: do your students understand the basics? Adv Physiol Educ *31*, 82-92.

Holmes D, Moody P, Dine D (2006). Research methods for the biosciences. Oxford, UK: Oxford University Press.

Kanari Z, Millar R (2004). Reasoning from data: How students collect and interpret data in science investigations. J Res Sci Teach *41*, 748-769.

Kardash CM (2000). Evaluation of an undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. J Educ Psychol *92*, 191–201.

Kish L (1965). Survey Sampling. New York: John Wiley and Sons, Inc.

Klahr D, Fay AL, Dunbar K (1993). Heuristics for scientific experimentation: A developmental study. Cognitive Psychol *25*, 111-146.

Kloser M, Brownell S, Shavelson R, Fukami T (2013). Effects of a research-based ecology lab course: a study of nonvolunteer achievement, self-confidence, and perception of lab course purpose. J Coll Sci Teach *42*, 72-81.

Koehler DJ (1994). Hypothesis generation and confidence in judgment. J Exp Psychol: Learn *20*, 461-469.

Kuhn D, Dean D (2005). Is developing scientific thinking all about learning to control variables? Psychol Sci *16*, 866-870.

Kuhn D, Pearsall S (2000). Developmental origins of scientific thinking. Journal of cognition and Development *1*, 113-129.

Kuhn D, Schauble L, Garcia-Mila M (1992). Cross-domain development of scientific reasoning. Cognition and Instruct *9*, 285-327.

Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics *33*, 159-74.

Laursen S, Hunter AB, Seymour E, Thiry H, Melton G (2010). *Undergraduate Research in the Sciences: Engaging Students in Real Science*. San Francisco: Jossey-Bass.

Lawson AE (1978). The development and validation of a classroom test of formal reasoning. J Res Sci Teach *15*, 11-24.

Lawson AE, Drake N, Johnson J, Kwon YJ, Scarpone C (2000). How good are students at testing alternative explanations of unseen entities? The American Biology Teacher *62*, 249-255.

Libarkin J, Ording G (2012). The utility of writing assignments in undergraduate bioscience. CBE-Life Sci Educ *11*, 39-46.

Lopatto, D (2003). The essential features of undergraduate research. Council on Undergraduate Research Quarterly *24*, 139-142.

Lopatto D (2004). Survey of Undergraduate Research Experiences (SURE): First findings. CBE-Life Sci Educ *3,* 27-277.

Lopatto D (2007). Undergraduate Research Experiences Support Science Career Decisions and Active Learning. CBE Life Sci Educ *6,* 297–306.

Lopatto D (2008). Exploring the benefits of undergraduate research: The SURE survey. In R. Taraban and R.L. Blanton (Eds.), Creating Effective Undergraduate Research Programs in Science. Teacher's College Press, NY. Pp. 112-132.

Metz AM (2008). Teaching statistics in biology: using inquiry-based learning to strengthen understanding of statistical analysis in biology laboratory courses. CBE-Life Sci Educ *7*, 317-326.

National Research Council (2007). Rising above the gathering storm: energizing and employing America for a brighter economic future. National Academies Press, Washington, DC.

NIST/SEMATECH (2003). Engineering statistics handbook. http://www.itl.nist.gov/div898/handbook/ppc/section1/ppc136.htm (accessed December 20, 2013)

Park J, Pak S (1997). Students' responses to experimental evidence based on perceptions of causality and availability of evidence. J Res Sci Teach *34*, 57-67.

Picone C, Rhode J, Hyatt L, Parshall T (2007). Assessing gains in undergraduate students' abilities to analyze graphical data. Teach Issues Exp Ecol 5 (July 2007): Research #1. http://www.esa.org/tiee/vol/v5/research/picone/article.html (accessed December 20, 2013)

Quinn GGP, Keough MJ (2002). Experimental design and data analysis for biologists. Cambridge University Press.

Ramsey FL, Schafer DW (2012). The statistical sleuth: a course in methods of data analysis. Brooks/Cole Publishing Company.

Roth WM, McGinn M K, Bowen G M (1998). How prepared are preservice teachers to teach scientific inquiry? Levels of performance in scientific representation practices. J Sci Teach Educ, *9*, 25-48.

Rubin DB (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics *29*, 185-203.

Ruxton GD, Colegrave N (2006). Experimental design for the life sciences. Oxford: Oxford University Press.

Salangam J (2007). The impact of a prelaboratory discussion on non-biology majors abilities to plan scientific inquiry. California State University, Fullerton. ProQuest Dissertations and Theses.

Schauble L (1996). The development of scientific reasoning in knowledge-rich contexts. Dev Psycho *32*, 102–119.

Schauble L, Glaser R (1990). Scientific thinking in children and adults. Contrib Hum Dev *21*, 9–27.

Schönborn KJ, Anderson TR (2009). A model of factors determining students' ability to interpret external representations in biochemistry. Int J Sci Educ *31*, 193-232.

Shi J, Power JM, Klymkowsky MW (2011). Revealing student thinking about experimental design and the roles of control experiments. Int J Sch Teach Learn *5*, 1-16.

SRI International, (2003). http://pals.sri.com/tasks/9-12/Testdrug/.PALS: Diffusion and Testing a New Drug 1. http://pals.sri.com (accessed 20 December 2013).

Singer, S.; Hilton, M.; Schweingruber, H. (2006). America's Lab Report: Investigations in High School Science; Washington, DC: The National Academies Press.

Sirum K, Humburg J (2011). The Experimental Design Ability Test (EDAT). Bioscene *37*, 8-16.

Sternberg RJ (2003). What is an "expert student?" Educ Res *32*, 5-9.

Thiry H, Weston TJ, Laursen SL, Hunter AB (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. CBE Life Sci Educ *11*, 260-272.

Timmerman B, Strickland DC, Johnson RL, Payne JR (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. Assessment and Evaluation in Higher Education *36*, 509–547.

Tobin KG, Capie W (1981a). The development and validation of a pencil and paper test of logical thinking. Educational and Psychological Measurement *41*, 413-424.

Tobin KG, Capie W (1982). Relationships between classroom process variables and middle-school science achievement. J Educ Psychol *74*, 441.

Wei CA, Woodin T (2011). Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE Life Sci Educ *10*, 123–131.

Wuensch KL (2001). Correlation and Causation http://core.ecu.edu/psyc/wuenschk/stathelp/Correlation-Causation.htm (accessed December 20, 2013).

CHAPTER 3: DESIGN AND DEVELOPMENT OF A 'NEURON ASSESSMENT' FOR MEASURING BIOLOGY STUDENTS' UNDERSTANDING OF EXPERIMENTAL DESIGN.

3.1 <u>Abstract</u>

Understanding breakthroughs in biology research and its future implications is important for undergraduate students to develop a correct impression of the source of knowledge in biology. There is need for students to develop abilities like designing experiments to generate evidence to pursue scientific questions relevant to them. This study describes the design and application of a new assessment, the 'Neuron Assessment' which examines whether undergraduate biology students are able to apply knowledge of experimental design to current research. Evidence from written responses followed by multi-phase oral interviews enables diagnosis of students' experimental design difficulties based on the Rubric for Experimental Design (RED) by Dasgupta et al. (2014). Furthermore, this paper uses the CRM model to examine the knowledge of experimental concepts (RC abilities) or representation for an experiment probed by the 'Neuron Assessment'. Findings indicate that experts and students reveal knowledge of a range of visual abilities and reasoning with concepts of experimentation when probed by the 'Neuron Assessment' which was missing before exposure to the assessment.

3.2 <u>Introduction</u>

Vision and Change (AAAS, 2011) listed formal practices like observation, experimentation and hypothesis testing among core competencies for disciplinary practice. These processes require students to understand how experimental design is performed in order to generate information about complex biological phenomena. In conversations with scientists, other research showed that when scientists explain

biological mechanisms, they construct experiments, they graph data, draw models of ideas they want to test and they also depict cellular and sub cellular locations. These approaches used by scientists were outlined in the MACH model (methods, analogies, context explaining how mechanisms work) of Trujillo et al., *in press*. In my previous research, by looking at responses to three published assessments, five key areas of experimental design knowledge were detailed in a Rubric for Experimental Design (RED). While published assessments helped us reveal major difficulties undergraduate students have with experiments, they did not carry probes to examine visual representation abilities such as those used by the scientists in the MACH study. Since scientists use diagrams to convey data from experiments when they explain biological mechanisms for MACH, we realized the need to design a question that both provides students with visual representations and allows them to generate their own visualizations when designing an experiment.

The MACH also highlighted the research context, hence a 'Neuron Assessment' was designed to understand how scientists and students approach reasoning about experiments using published visualizations and representations they create for themselves when they design experiments on isolate neurons to answer questions about a disease as a current research context. This chapter characterizes the usefulness and limitations of the 'Neuron Assessment' for revealing expert and students' thinking about experimental design concepts and diagrams in the context of a human disease that might be understood and explained by experimenting with the function of biological molecules in a neuron cell.

3.3 Background

Previous work reveals that undergraduate students face challenges with aspects of experimental design like knowledge about the experimental subject (Salangam, 2007), manipulating variables (Picone et al., 2007, Shi et al., 2010), identifying measurable experimental outcomes (Hiebert, 2006; Harker, 2009), recognizing sources of variation (Kanari & Millar, 2004; Kuhn & Dean, 2005) and drawing causal inferences (Klahr, Fay & Dunbar, 1993; Schauble, 1996). We designed a Rubric for Experimental Design (RED)

(Dasgupta et al., 2014) to characterize five broad areas of students' experimental design difficulties: a) variable property of an experimental subject, b) manipulation of variables, c) measurement of experimental outcome, d) accounting for variability, and e) scope of inference of findings. Difficulties in these areas were detected in student responses to published assessments. The 'Shrimp Assessment' presents a context where students manipulate various growth enhancing nutrients and salt levels to design an experiment to track growth of tiger shrimp. The 'Drug Assessment' examines abilities to design an experiment to test a blood pressure drug.

Schönborn & Anderson's (2009) CRM model proposes that engagement in any kind of scientific thinking requires interactions among three factors: conceptual knowledge (C), reasoning skills (R) and mode of representations or visualizations (M). Factor CM or concepts and the mode of representing them involve conventions used by scientists when they visualize an experiment. Various skills are involved in recognizing and creating visual representations (Schönborn and Anderson, 2009) like decoding the symbolic language and interpreting and using the representations when creating your own graphs. More complex visualization skills include horizontal translation across alternate representations of the same biological phenomenon and visualizing levels of organization from an organism to the level of a cell or molecules relevant to biological phenomena. These visualization skills (RM abilities according to CRM) are required for scientists to interpret and design experiments and thus our rationale was to evaluate if these skills that experts apply are also applied by students. Similarly, describing the design of a hypothetical experiment requires application of knowledge of the concepts relevant to the subject matter and also experimental design concepts (RC abilities according to CRM). Therefore, in this study we examine and compare knowledge of concepts that experts and students present as they propose an experiment using the subject matter of the 'Neuron Assessment' as context. A glossary of experimental design concepts (Dasgupta et al., 2014) was used as a guide to identify concepts presented by experts and students in their explanations.

In the context of neuron functions, factor CM or conventional modes of representing mitochondrial transport along axons would involve globular or spherical shaped mitochondria moving along elongated rod like axons as shown by experts and textbook images. Similarly, conventional ways of representing an experiment would be graphical representations of data with the dependent variable on the y-axis to display experimental findings. Factor, RM indicates reasoning about the modes of representations or visualizations used to represent experimental ideas. For example, reasoning about graphical representations involves organizing the treatment and outcome variables appropriately on the x- and y-axes. Factor RC refers to reasoning about the concepts related to experimental design, for example, reasoning about treatment and outcome variables to show presence or absence of a causal association in an experiment.

In previous work with the RED, student difficulties with experimental design were only characterized for the RC category because the assessments used to develop RED did not include any diagrams and students were not prompted to create any visual representation of experiments. Thus, CM or RM abilities such as construction of graphical representations or reasoning about experimental variables using a graph were not examined. The current study builds on previous work by exploring how students use visualizations when they design experiments.

For the current study, the CRM model was used to guide the design of an original assessment in the context of a cutting edge research problem. The assessment was designed to provide students with information about transport of mitochondria in cells with supportive diagrams. Providing students with necessary subject matter knowledge would allow us to focus on their experimental design abilities while the diagrams would provide insight into how well students interpret and represent visual information that experts or textbooks use to depict transport of mitochondria in cells. The research problem posed by the 'Neuron Assessment' asks for a method to investigate the source of a disorder associated with mitochondrial movement along axons in neurons. The CRM model is a useful tool to characterize how experimental design is represented through

visual modes when faced with designing an experiment to address a cutting edge research problem.

This study examines the usefulness of the 'Neuron Assessment' to compare expert and undergraduate student knowledge about experimental design. The overall goal was to use the assessment to probe for expert ways of designing an experiment and to validate if the question was useful to discriminate novice answers from more expert responses. The study addresses two research questions:

RQ1. How well does the 'Neuron Assessment' reveal the nature of expert knowledge about organelle movement in neurons and the experiments used to elucidate that knowledge?

RQ2. How well does the 'Neuron Assessment' expose student knowledge and related difficulties with experiments to investigate organelle movement in neurons?

To get a deep understanding of differences in how students and experts think about experiments, a case study method was used to answer these questions. Case studies allow exploration of situations in which the intervention has no preconceived set of outcomes but rather involves examining expert and student knowledge and visual representations of experimental evidence without any relevant behaviors being manipulated. It also covers contextual conditions and allows understanding of the underlying participant experiences and how they influence outcomes from the study (Yin, 1984).

If the 'Neuron Assessment' can be demonstrated to be a useful measure for discriminating different levels of understanding of experimental design, we expect it will provide an opportunity for experts as well as students to present their knowledge and visual depictions related to experiments regardless of their prior knowledge of the subject matter related to neurons and the movement of mitochondria in neurons.

3.4 <u>Method</u>

To understand the usefulness of the 'Neuron Assessment' as a probe to reveal expert (RQ1) as well as students' knowledge (RQ2) about experimental design, we initially designed and piloted the 'Neuron Assessment'. The assessment format was modified to provide clear background information and to minimize any confusion. A neuroscientist was recruited as an expert research participant in the experimental design case study with an oral interview to examine the potential of the 'Neuron Assessment' to reveal the nature of expert knowledge about experimental design concepts and visualizations (RQ1). Then student interviews were conducted and analyzed for presence of difficulties with experimental concepts and visuals using expert responses as comparison (RQ2) and RED as a tool to characterize expected difficulties. Each of these steps is detailed in the following sections.
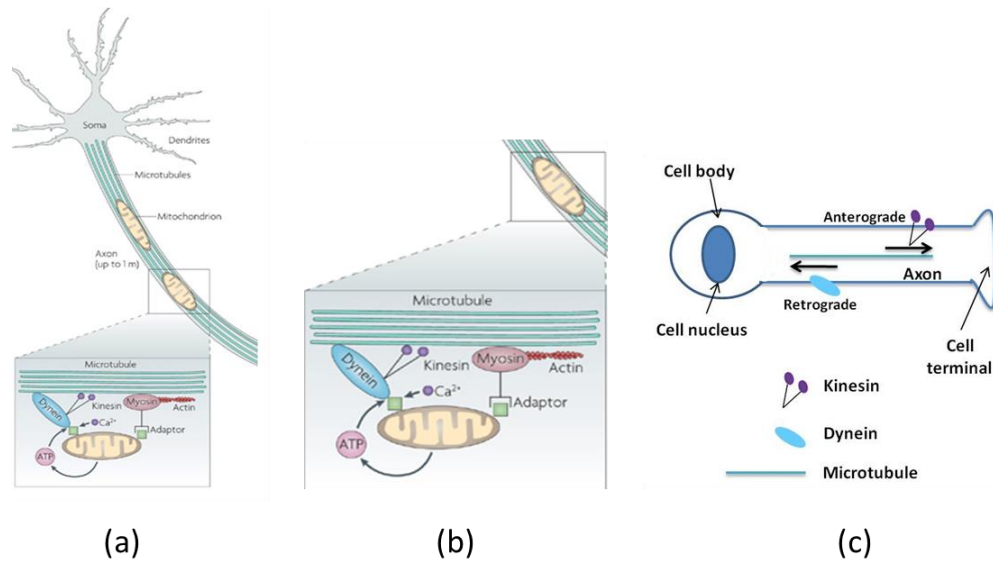
(a)   (b)   (c)

Background: Mitochondria are one of the several organelles that get transported across the axon of a nerve *(Refer to figure above)*. They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions: Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder associated with impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons. The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;   **Compound D:** inhibits dynein;
Imaging software: measures mitochondrial movement in neurons.

a) Describe what you see in the three diagrams above. Please share in detail what you think about it.
b) What could be a potential hypothesis for your experiment? Create a representation to illustrate your hypothesis.
c) Which factors will you vary and which will you keep the same in your study? Why? Use a visual representation to illustrate the factors you will vary or keep same.
d) How will you assign subjects to groups for your experimental study? Explain. Create a representation to support your answer.
e) Do you think you can establish a cause-and-effect relationship between the treatment and a response variable in this experiment? Justify your answer. Create a visual representation to illustrate a cause and effect relationship.
f) How would you present the results of your experiment?
g) What results do you expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of your experiment. What visual representation will allow you to present results?
h) How will you improve the validity of your experiment? What visual representation will you use to show how the validity will be improved?
i) What do you think this diagram is not showing? Explain your answer.
j) Is there anything about this question that you don't understand or find confusing? Explain.
k) Consider yourself a diagram designer. If you could change the diagrams, what would you change or how would you improve them?

Figure 3.1: The 'Neuron Assessment' background information and supporting figures.

3.4.1 Design of the 'Neuron Assessment'

The 'Neuron Assessment' prompts design of an experiment to investigate about a disorder related to organelle movement in neurons (Figure 3.1). Each part of Figure 3.1(a-c) was logically organized to represent complementary perspectives of organelle movement along neurons based on visual design principles as recommended by Mayer and Moreno (2003). Background information and the diagrams were provided to level any differences in students' prior subject matter knowledge in order to assess knowledge of experiments, rather than cell biology. Visual representations have been shown to alleviate misinterpretation by translating across multiple modalities (Stenning and Oberlander, 1995; Mayer and Moreno, 1999). The 'Neuron Assessment' was designed with written probes to diagnose understanding in each of the five RED areas. To probe understand of experimental subjects, the assessment (Figure 3.1) asks, *"How will you assign subjects to groups for your experimental study? Explain;"* to probe for knowledge of treatment/control conditions, the prompt asks, *"Which factors will you vary and which will you keep the same in your study? Explain why;"* to probe for understanding of the questions, *"How would you present the results of your experiment?"* and *"Do you think you can establish a cause-and-effect relationship between the treatment and a response variable in this experiment? Justify your answer"* probe for knowledge of measurable outcomes; the assessment probes abilities for dealing with variation and interpreting and representing experimental ideas by asking, *"How will you improve the validity of your experiment?"* and *"What results do you expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of your experiment"*. Once designed, the assessment was piloted with a small sample of first year undergraduates as the intended study population.

3.4.2 Piloting the 'Neuron Assessment'

Two sessions were conducted in Fall 2010 and Spring 2012 to pilot the 'Neuron Assessment'. In 2010, 18 first year undergraduate students and three advanced students (two graduate students and one advanced undergraduate student) participated as volunteers. The assessment was administered as a 2-tier multiple-choice test in paper-

pencil format. Analysis of responses showed that the 2-tier format provided only limited information on the nature of students' problems with designing experiments. Therefore, a second pilot was conducted with a modified open ended version of the assessment which was also administered as a paper pencil test. Five experts (one faculty member, two graduate students and two advanced undergraduate students) and 15 first year undergraduates participated as volunteers. The pilot study was followed by interviews of the participants who clarified how the 'Neuron Assessment' could be modified to probe for the five RED areas. This second pilot study also revealed that some students used drawing to explain their ideas about experiments and so the probes were modified to prompt for drawings to illustrate the role of visualization in designing experiments (Figure 3.1).

3.4.3 Research Participants

Prior to the study, research procedures were reviewed and approved by the Institutional Review Board (IRB) and written consent forms were filled out by each participant (Appendix P). Upon finalizing the assessment, a scientist who studies neurobiology was recruited as the "expert" volunteer. The expert's research area of focus was related to but did not directly involve the same context as the story of mitochondria movement for the 'Neuron Assessment'.

Student volunteers were recruited from a first year undergraduate introductory biology course (Biology II: Development, Structure, and Function of Organisms). This course was appropriate because a key learning objective was to gain biology knowledge through evidence from research and experimental design and also to practice drawing graphs to represent findings. In 2013, at the beginning of the semester before any material dealing with experimental design was covered, as a normal part of their class, students completed a survey via Qualtrics® online survey software. The survey offered a sign up opportunity to all enrolled students to participate in the experimental design activity. Thirteen students agreed to participate.

Using a purposeful sampling strategy, four students were selected for this study. The selection was based on following criteria: each student was at the first year undergraduate level, each student represented a different major, and subjects were selected for broad representation of gender and ethnicities. Prior knowledge or ability was not a factor known to the instructor when these students were recruited at the beginning of the semester but these students were identified by the instructor as verbally expressive and capable of sharing their own ideas with clarity. The student participants are identified with pseudonyms Juan, Daniel, Eve, and Li Na for confidentiality. The expert is referred to as Eric. Juan is a male Hispanic who is a chemistry major. Eve is a Caucasian female and microbiology major. Li Na is an Asian female who majors in cell and molecular biology. Daniel is a Caucasian male and engineering major. The expert is a Caucasian, male neurobiology research scientist.

3.4.4 Study Procedure

The written experimental design activity was completed within one hour by each participant. Then a follow-up oral interview lasted on average two hours immediately after the written session. Oral interviews were recorded digitally and transcribed. On average, each interview involved six hours of transcription. Data files were stored on a secured computer, and files were transferred using a secure, password protected file transfer system as per IRB protocol #1008009581.

3.4.4.1 The Three Phase Seated Interview Technique Format

The three phase seated interview technique (3P SIT) from Schönborn & Anderson (2009) was adapted to include an initial phase (phase 1) with questions to understand each participants knowledge of concepts related to mitochondrial transport in neurons and experimental design before exposing them to the 'Neuron Instrument.' For example, questions asked were "*What comes to mind when I say 'neurons'?*" or "*What comes to mind when I say 'organelle movement'? Please draw to help me understand what you mean.*" In the next phase (phase 2) participants were provided with the 'Neuron Assessment' to study the impact of the visuals and background information and further examine how they present their knowledge of experimental design when faced with a

current research problem. To understand if the story and diagrams about transport of mitochondria was intelligible and to find out if the 'Neuron Assessment' was clear enough to expose their thinking about experimental design, a third set of questions (phase 3) was asked to gather reflections on phase 1 and 2.

3.4.4.2 CRM Coding of Interview Responses

The CRM coding method of Schönborn and Anderson (2009) was applied to analyze the data. This involved inductively examining the data to code information into three categories, CM, RM and RC. First, CM or expert conceptual knowledge depicted by the mode of representations deployed by the expert was identified. The expert drawings were examined to identify parts that depict conventional modes of representing both experimental and biological concepts related to neurons and organelle movement using visuals and associated symbolism. CM abilities was added to the Glossary (Appendix G) and RED (Appendix I). To identify knowledge presented, we modified the RED to include 'propositional statements' corresponding to visual representations for RED components. Further, our original glossary list of vocabulary terms associated with each of five RED areas (Dasgupta et al., 2014) was modified to include how experimental concepts are visualized (Appendix G). The second category, RM or reasoning with mode of representations involved inductively identifying the data that indicates reasoning with specific representations. The third category or RC indicates retrieving or reasoning with their conceptual knowledge of biology subject matter and experimental design concepts in their design of an original experiment. The expert responses were examined to look for parts of an experiment depicted in the form of visuals. This information was added to the glossary and thus, the glossary list was modified and used as a guide to examine visual modes of parts of an experiment presented by students. This list was subsequently validated using the analysis of the visual data provided by the expert and students (Column 1, Table 3.12).

To answer the first research question about how well the 'Neuron Assessment' reveals the nature of expert knowledge about organelle movement in neurons and the experiments used to elucidate that knowledge the expert 3PSIT interview responses

(Appendix K) were transcribed and analyzed using the CRM framework. The transcript and associated drawings were examined for the conventions used to describe mitochondrial transport and these conventions are listed in a Table 3.1. The various visual abilities demonstrated by the expert as he reasoned with diagrams (RM abilities) to represent mitochondrial movement in neurons and experimental design, both before and with the 'Neuron Assessment' were analyzed. These findings are organized into another table for easy comparison (Table 3.2). Finally to compare how the expert reasoned about concepts before and with the 'Neuron Assessment', the expert interview was coded for knowledge of concepts (RC) relevant to mitochondrial movement and for each component of the RED. The glossary list (Appendix G) was referred to determine correct knowledge of the experimental concepts presented by the expert. The RC abilities were organized into Table 3.3 to show specific underlying concepts the expert used related to each of the RED components. For example, Table 3.3 compares how the expert reasoned with an underlying concept related to the RED component, *variable property of the experimental subject*, before and with the 'Neuron Assessment'.

To answer research question *2* about *how well does the 'Neuron Assessment' expose student knowledge and related difficulties with experiments to investigate organelle movement in neurons*, the student 3PSIT interviews were transcribed and analyzed using CRM coding for each of the four student participants: Juan, Li Na, Eve and Daniel. The interview transcripts (See raw interview in Appendix K) were subjected to inductive coding using RED to diagnose students' knowledge of and difficulties with diagrams and concepts for the design of an experiment both before and with the 'Neuron Assessment'. Tables 3.4-3.7 were generated to compare diagrams student created before and with the 'Neuron Assessment'. These were analyzed to identify correct knowledge and difficulties with R-M abilities pertaining to mitochondrial movement in axons and to experimental design. Tables 3.8-3.11 were generated to compare how well each student performed before and with the 'Neuron Assessment' on concepts related to mitochondrial movement in neurons and each component of RED as they reasoned about their design of a hypothetical experiment.

3.5 Results

3.5.1 Expert Abilities Probed By the 'Neuron Assessment'

Findings highlight the nature of expert knowledge revealed before and with the 'Neuron Assessment' using the guiding CRM framework. In general, expert CM or conventional use of representations with the 'Neuron Assessment' includes neurons, organelles, motor proteins, microtubules, arrows to point out features and show movement, an experimental design table with treatment groups, and graphs (Table 3.1). Expert RM abilities displayed in Table 3.2 shows reasoning with diagrams and experimental design visualizations both before and the 'Neuron Assessment'. Finally, Table 3.3 compares how an experiment was designed using knowledge of specific experimentation concepts (RC) both before and with the 'Neuron Assessment'. Expert RM and RC abilities were characterized according to evidence of correct ideas (green cells) and for lack of evidence (orange) when any information for a certain RED component was missing.

3.5.1.1 Expert CM Abilities

Table 3.1 summarizes the conventional modes of representing concepts illustrated in Figure 3.2 when the expert depicted neuron components or parts of experimental design. The expert illustrated with diagrams several different conventional ways of presenting mitochondrial movement along axons (Figure 3.2A-C) and diagrams were drawn to show how information is organized for the design of experiments (Figure 3.2D-F). For example by convention, neurons are presented with a circular cell body and elongated axons (Table 3.1, top row), whereas experimental findings are represented using tables and graphs with various parts (Table 3.1, bottom row).

Table 3.1: Propositional knowledge presented by the expert with figures (CM)

CM	Conventions
Neurons	Circular cell body, elongated axons, small dendritic processes (Figure 3.2A)
Organelles	Globular (Figure 3.2B and 3.2C)
Motor proteins (kinesin and dynein)	Stick figure (Figure 3.2B)
Microtubules	Long strands (Figure 3.2B-C)
Arrows to identify components	Points at features, movement in anterograde and retrograde directions (Figure 3.2A-B, D and F)
Arrows to show movement	Points at features (Figure 3.2B)
Experimental design table	Control and treatment group variables organized into separate columns (Figure 3.2E)
Graph	Independent variable on x-axis, dependent variable on y-axis, key to symbols on the graph show measures for treatment and control groups depicted as separate points or separate bars (Figure 3.2F).
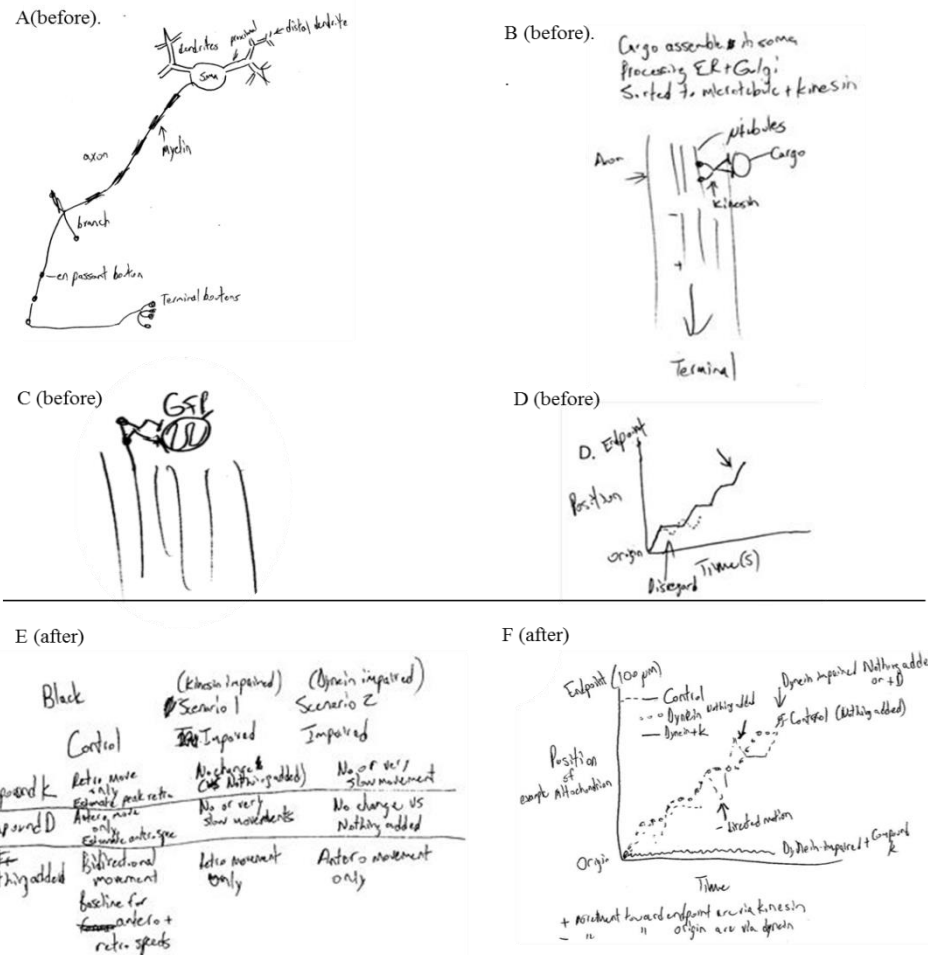
Figure 3.2: Expert's 'Neuron Assessment' figures

Before 'Neuron Assessment': A. Neuron concepts: Spatially manipulate provided 'Neuron Assessment' figures to interpret and explain a concept of a neuron. Visualize levels of organization, relative size, shape and scale of cell body and axon. B. Organelle movement in neurons: Use representations to interpret temporal resolution of steps in cargo transportation along microtubules during vesicular/organelle transport across neurons. Translate horizontally across multiple representations of various aspects of mitochondrial movement. C. Interpret and use a representation (provided neuron figures) to demonstrate design of an observational experiment (GFP labeled tracking of mitochondria). Construct a representation to suggest an observational experiment (GFP labeled tracking of mitochondrial movement along neurons). Note that experimental treatment groups were not indicated. D. Interpret and use provided neuron figures to demonstrate design of an observational experiment to track GFP labeled mitochondria). Construct a graph to represent findings from GFP labeled tracking of mitochondria with independent variables and dependent variables on x- and y-axes respectively. Specific treatments are represented as curves. Dotted line presents outliers as a result of variation. Translate horizontally across multiple figures of mitochondrial movement. Interpret the temporal resolution of mitochondrial movement along neurons – position of organelle along axon over time. With 'Neuron Assessment': E. Neuron concepts: Decode the symbolic language composing neurons in 'Neuron Assessment' figures 3.1a-c. Translate horizontally across multiple representations of neurons. Organelle movement: No additional figures were drawn to show organelle movement. E. RED Areas: Interpret provided neuron visuals to design experimental treatment groups. Construct experimental groups to represent manipulation of control and treatment variables. Interpret and use a representation (neuron figures) to solve a problem (investigation of organelle movement in neurons). F. Construct a graph to represent curves corresponding to control and treatment outcomes. Construct a graphical representation with independent variables and dependent variables on x-axis and y-axis respectively. Different treatments are represented as separate lines. Translate horizontally across experimental table and experimental graph with each treatment as a separate curve. Interpret the temporal resolution of mitochondria movement along neurons – position of organelle along axon over time.

3.5.1.2 <u>Expert RM Abilities</u>

Table 3.2 compares how the expert reasoned during the interview modes of representing information (RM) before and with the 'Neuron Assessment'. The expert both created visuals as well as used those provided when he reasoned about neuron functions and experimental design (RM). Figure 3.2 (A-F) shows the expert's showcases visual representations that together with the quotes from the interview (Appendix K) provide evidence for the abilities listed in Table 3.2.

Before seeing the 'Neuron Assessment', the expert produced diagrams of a neuron (Figure 3.2A), mitochondrial movement (Figure 3.2B) and depicted tracking of labeled mitochondria (Figure 3.2C and 2D) but illustrated no experimental groups. However with the 'Neuron Assessment', the expert provided figures and demonstrated RM abilities with experimental tables and graphs (Figure 3.2F-G) relevant to all five RED components (Variable property of an experimental subject, Manipulation of variables, Measurement of outcome, Accounting for variability and Scope of inference). Thus, the expert visualized components of an experiment better with the assessment than before being prompted by the 'Neuron Assessment' questions.

Table 3.2: Experts' reasoning with visualizations (RM) before and with the 'Neuron Assessment'

Concepts	RM[*]	Before	With
		Neuron subject matter	
a. Neuron knowledge	**Spatially manipulate a representation** (Figure 3.2A)[#]	Spatially manipulate a representation (figure of a neuron) to interpret and explain a concept (neuron anatomy).	
	Visualize levels of organization	Visualize levels of organization, relative size and scale (relative size and shapes of cell body, axon and mitochondria).	
	Decode a representation		Decode the symbolic language composing a representation (Figure 3.1a-c)
	Translate horizontally across representations		Translate horizontally across multiple ERs of organelle movement in neurons (Figure 3.1a-c).
b. Organelle movement in neurons	**Interpret temporal resolution**	Temporal resolution of steps in cargo transportation along microtubules during cellular processes of vesicular/organelle transport across neurons (Figure 3.2 B)	Temporal resolution of mitochondria movement along neurons – position of organelle along axon over time (Figure 3.2F).
	Translate horizontally across representations	Translate horizontally across multiple ERs of a concept (multiple figures representing various aspects of organelle movement) (Figure 3.2B).	Translate horizontally across multiple representations of neurons (Figure 3.2E).
		RED areas	
c. Experimental design representations Control group[20θ] Treatment group[18]	**Interpret and use a representation**	Provided neuron figures were interpreted to demonstrate design of an observational experiment involving GFP labeled tracking of mitochondria (Figure 3.2C).	Provided neuron visuals were interpreted to design experimental groups and solve a problem of investigation of organelle movement in neurons (Figure 3.2F).
	Construct a representation	The representation suggests an observational experiment (GFP labeled tracking of mitochondrial movement along neurons) but no experimental groups were identified.	The representation represents manipulation of control and treatment variables organized as separate groups in a table (Figure 3.2F) [2, 18].
d. Graphs RED areas: Variable[20]	**Interpret and use a representation** (Figure 3.2C)	Provided neuron figures were interpreted to demonstrate design of an observational experiment involving GFP labeled tracking of mitochondria.	

---

[*] RM from the list in Table 3.1

[#] Numbers and letters in parentheses refer to the expert's diagram in Figure 3.2

[θ] Superscripts refer to the concepts listed in column 1 and defined in Appendix G

Table 3.2: Experts' reasoning with visualizations (RM) before and with the 'Neuron Assessment'

Concepts	RM[*]	Before	With
property of an experimental subject Manipulation of variables[17] Measurement of outcome[70] Accounting for variability[22] Scope of inference[15]	**Construct a representation**	Graph constructed to represent findings from the observational experiment) (Figure 3.2D) [3, 5]. Graph constructed with independent variables and dependent variables on x-axis and y-axis respectively. Specific treatments are represented as curves. Dotted line present outliers from variation (Figure 3.2D) [2, 3, 4].	Graph constructed to represent control and treatment variables organized as separate curves) (Figure 3.2F) [1, 2]. Graph constructed with independent variables and dependent variables on x and y-axes[2]. Different treatments are represented as separate points. Dotted line present outliers from variation[4] (Figure 3.2F).
	Translate horizontally across representations	Horizontal translation across multiple representation of an observational experiment tracking movement of mitochondria along axons (Figure 3.2C-D).	Horizontal translation across experimental table and experimental graph representing each treatment in the table as separate curves on the graph appropriately (Figure 3.2E-F).
	Interpret the temporal resolution of representations	Movement of organelle along neurons– position of organelle along axon over time depicted[3] (Figure 3.2D).	Movement of organelle along neurons– position of organelle along axon over time depicted (Figure 3.2F).

[*] Superscripts refer to the concepts listed in column 1 and defined in Appendix G

Based on the representational modes presented by the expert, the original glossary list by Dasgupta et al. (2014) was revised (Appendix G) to incorporate visual modes of representation for parts of experimental design. Definitions for visual representation for a control (Appendix G, Term 1), cause and effect relationship (Term 4), factors (Term 5), outcome variable (Term 7), sample size (Term 14), subject (Term 16), treatment variable (Term 17) and variability (Term 22) were included. Consequently, the RED was also modified to incorporate visual evidence associated with each RED area (Appendix I) as detailed in the next paragraph.

The expert depicted control and treatment variables in the experimental table (Appendix G, term 1; Appendix I, RED area 1) and as curves on the x-axis of his graph (Figure 3.2F). Experimental factors were identified in the graph figure legend (Appendix G, term 5; Appendix I, RED area 2). Outcome variables and causal relationships could be interpreted from graphs x- and y-axes labels (Appendix G, term 5 and 7; RED area 3 and 5). The expert showed variation with tracking of position of a mitochondrion and thus ways to represent variability in a graph were added (Appendix G, term 22, RED area 4).

The expert figures highlighted modes of representation as he drew when designing an investigation of mitochondrial movement. The expert decoded neuron knowledge presented in symbols (Table 3.2, row a). He used the provided figures and constructed ones of his own to design an experiment (Table 3.2, row c and d). He used alternative representations to present knowledge of the organelle movement and thus showed horizontal translation (Table 3.2; row a and b). Neuron structure was illustrated from organelle to cellular levels (Table 3.2; row a). Neuron anatomy was also spatially manipulated to explain various parts of an experiment (Table 3.2, row a)

3.5.1.3 Expert RC Abilities

Table 3.3 shows that the expert used concepts related to the neuron subject matter as well as experimental design concepts when explaining experimental evidence both before and when exposed to the 'Neuron Assessment'. A superscript number for each concept corresponds to the glossary list (Appendix G). R-C abilities in adjacent columns

show what the expert did or how the concept was used at each stage of the interview. Evidence was identified either when the participant used the specific term or provided an explanation that indicated knowledge of the concept as defined in the glossary. For example, evidence of knowledge about 'variability' using replication was marked as present when the participant stated 'replicate the treatments to consider variability among outcomes' or 'repeat the treatments to obtain a range of values for the same outcomes'.

Table 3.3: Experts' reasoning with experimental design concepts (RC) before and with 'Neuron Assessment'.

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
			Neuron subject matter		
(I) Neuron concepts	a. Neuron knowledge	i. *"[In a neuron] there would be dendrites, an axon which can be myelinated, circular soma and some dendritic branches going up."*	Memorize entities: axon, dendrites, myelination, soma	ii. *The dendrites and an axon are typically parts of a neuron.*	Memorize entities: axon, dendrites
	b. Organelle movement	i. *"[In organelle movement] the cargo is sorted to microtubules and kinesin. So we have microtubules bundles going down the axon and then the kinesin heavy chain help in transporting the cargo (could be organelles) across an axon in a neuron."*	Apply knowledge of concepts (molecules like kinesin, microtubules, kinesin heavy chain) to explain organelle movement	ii. *"In this study there are trying to test the mechanism for a particular set of neurons with impaired mitochondrial movement, to figure out how to correct the impairment and apply that to repair or preventing of neurons in patients with the disorder. They are already down to the idea that a defect with either kinesin or dynein is causing the disorder."*	Apply knowledge of concepts (neurons, molecules like kinesin, microtubules, dynein) to explain investigation goal of diagnosing impaired mitochondrial movement.
			RED areas		
(1) Variable property of experimental subject	a. Experimental subject *Sample[13#] Subject[16] Unit[19] Variable[20]*	i. *"We have GFP-tagged mitochondria[16] and then we have microtubules[16] which will be attached to kinesin. Basically then we will use a fluorescent microscope to track (moving[20]) mitochondria[16]."*	Integrate knowledge of concepts (mitochondria, microtubules, kinesin, fluorescent microscope) with experimental subject[16] and its variable[20] property i.e. movement.	ii. *"We will do a position vs. time[20] of mitochondria and looking along the axons of neurons[16]. We will use neurons are derived from the cell cultures of neurons[16] of patients/cell lines with the impairment [13]. There will be scenario one with kinesin impaired and scenario two with dynein impaired neurons[19]"*	Apply knowledge of concepts (neuron cell cultures) to propose an experimental subject[16]along with a variable[20] property (impairment).

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.3: Experts' reasoning with experimental design concepts (RC) before and with 'Neuron Assessment'.

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
(2) Manipulation of variables	a. Treatment variable *Subject[16] # Variable[20] Treatment variable[17] Treatment group[18]*	i. *"Using live cell imaging and a fluorescent tag to tag some mitochondrial specific protein and track fluorescence as it moves down the axon."*	Lack of Evidence	ii. *"To each of these kinesin impaired and dynein impaired cell lines[18]. I will add compound K, compound D respectively as treatments[17]"*	Transfer and apply knowledge of variable[20] property of the experimental subject[16] (kinesin/dynein impaired neurons) to propose treatment (independent) variables[17] (compound K/compound D).
	b. Control variable *Control[1] Control group[2]*	i. *"I am guessing since we are only tracking movement in the neurons, a control [1, 2] won't be necessary at this point."*	Lack of Evidence	ii. *"We will have a control (normal neurons[1]). When nothing is added, we get baseline for anterograde/retrograde speed. To a group of normal neurons we will add compound K and D respectively.[2]"*	Transfer and apply knowledge of the concept of control[1, 2] for comparison purposes.
	c. Controlling outside variables *Confounding variables[8] Control group[2] Treatment group[18] Variation[21]*	i. *"The axons in the study obviously should be picked from the same kind of neurons[21] to avoid confounding factors[8] that might contaminate our findings."*	Apply knowledge of ways to reduce variation[21] by controlling confounding variables[8].	ii. *"The factors [across treatment[18] and control group][2] kept the same would be the imaging set up, conditions of the medium, the cell culture age, time window used to measure, effective concentrations of the inhibitors etc[8]. This ensures that any external sources of variation[21]n are removed in the experiment."*	Apply knowledge of matching treatment[18] and control group[2] variables to propose ways to deal with variation[21] from confounding variables[8].

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.3: Experts' reasoning with experimental design concepts (RC) before and with 'Neuron Assessment'.

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
(3) Measurement of outcome	a. OutcomeVariable[20] *Subject[16] Outcome variable[7]*	i. *"We then quantify the movement of the particle[7, 16, 20] along a certain segment of axon."*	Apply knowledge of variable[20] property of experimental subject[16] under investigation to propose measureable outcome variables (movement of particles). Reason locally about outcome variables[7] (movement of particles along the axon).	ii. *"So in a control cell from normal patients, both anterograde and retrograde movement will take place towards the end point (100 μm). In the same kind of cell from normal patients, when compound D is added, we will notice anterograde movement only in the positive direction (100 μm)[7]. What we observe in the normal cells upon treatment with inhibitors can be then compared with the cells from the patients with the disease to test what we find in our study actually applies to the real patients."*	Apply knowledge of variable property of experimental subject[16] (anterograde/retrograde movement) under investigation to propose measureable outcome variables[7] (movement of particles).
(4) Accounting for variability	a. Replication[12#] *Variability[22] Subjects[16] Units[19] Treatment group[18] Control group[2]*	i. *"We will be using multiple neurons[16, 19] and using the method I described, we can obtain several values[12] for the speed of mitochondria moving towards an end point in the selected field which can be averaged[22] eventually."*	Apply knowledge of ways to reduce variability[22] from experimental subjects[16] or units[19] by averaging values as a result of replication[12].	ii. *"We would take measurements [for the treatment and control groups] multiple times[18]. Even though we think we have similar cells[16, 19] and conditions, there is going to be some variability[22] between them and we want to determine the extent of variability[16]"*	Apply knowledge of ways to measure and reduce variability[22] by replicating[12] measurements on multiple cells[16, 19] in treatment[18] and control groups[2].
	b. Randomization[1] *Random sample[10] Treatment groups[18],*	i. *"We will be using multiple neurons picked randomly[10] and then set up probably assigning sets of neurons[18] in a randomized manner[11] to several petri-dishes."*	Apply knowledge of ways to reduce variation[22] by randomized assignment[11] of treatments[18].	ii. *"Randomly assigning[11] cells [of blind origin][10] to 3 [treatment] groups[18] reduces bias during the experiment and accounts for variability among measures[22]"*	Apply knowledge of ways to reduce variability[22] by selecting a random sample[10] and by randomization[11] of

Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.3: Experts' reasoning with experimental design concepts (RC) before and with 'Neuron Assessment'.

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
	Variability [22]				treatments[18].
	c. Representative sample[10] *Sample* [13] *Random sample* [10] *Control group* [2] *Treatment group* [18]	i. *"Often in textbook, the spinal motor neurons are shown as the representative neurons[10] but they are not really representative of all kinds of neurons in the brain with a big fat axon and sparse dendrites[13]. That's probably not true for 90% of neurons."*	Memorize knowledge of spinal motor neurons[13]. Apply knowledge of representative sample [10, 13](of neurons) as measure to account for variation.	ii. *"I would be blind as to the origin of the cell [10, 13]-so they wouldn't know whether the representative neurons are derived from the patient population (treatment group) [18] or the normal human cell line (control group)[2]"*	Transfer and apply knowledge of representative sample[10] (of neurons) to sample[13] of experimental subjects as part of treatment group [18].
(5) Scope of inference	a. Scope of Inference[15#]	i. *Our goal was to measure organelle movement within the axon. We fluorescently labeled particular organelle-mitochondria along the axon and then tracked its motion using live cell microscopy. We quantified those movements by looking at multiple sets of neurons to determine the positions of mitochondria and determined velocity and see whether there are different forms of movement.*	Lack of evidence	ii. *"What we observe in the normal cells upon treatment with inhibitors can be then compared with the cells from the patients with the disease to test what we find in our study actually applies to the real patients[15]"*	Reason locally and globally about scope of inference[15] to make conclusions about an investigation.
	b. Cause and effect[4] *Treatment variable* [17] *Control variable* [1] *Outcome variable* [7] *Correlations* [3]			ii. *"We might take a patient with the disorder[17], and because we know that most probably the patient has dynein impairment, when we add compound K (inhibits anterograde movement), we will see zero to no movement."[7] "The conclusion from this graph is that the dynein*	Apply knowledge of treatment[17], control[1]and outcome[7] variables to develop causal[4] explanations.

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.3: Experts' reasoning with experimental design concepts (RC) before and with 'Neuron Assessment'.

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
				is impaired because in the control we see some proportion of retrograde motion but with dynein impaired we see only movement in the positive direction/anterograde movement.''[3,4]	

■ Correct Ideas ■ Difficulties ■ Lack of evidence

Before the 'Neuron Assessment' (Phase 1), the expert demonstrated knowledge of neuron concepts but did not propose an experiment with a control group for comparison to test organelle movement in neurons. When the expert said, *"Using live cell imaging and a fluorescent tag to tag some mitochondrial specific protein and track fluorescence as it moves down the axon"*, this revealed an observation with no experimental treatment variables. However with the 'Neuron Assessment' (phase 2), the expert said *"To each of these kinesin impaired and dynein impaired cell lines I will add compound K, compound D respectively as treatments".* This demonstrates an experimental intervention with treatment variables. During Phase 3 the expert said*, "I think this is a fairly clear question. You can set up the experiment in a way that will give you some form of answer so it does lead you to derive a certain answer if you have the right ideas about designing an experiment".* These findings indicate that the 'Neuron Assessment' carried sufficient information to design an experiment to experimentally investigate organelle movement in neurons.

In summary, analysis of the expert response to the 'Neuron Assessment' demonstrated that the assessment was useful to probe knowledge about neurons and organelle movement in neurons and the item was effective at revealing the experimental design components identified in the RED. Since the 'Neuron Assessment' was valid for revealing expert knowledge of experimental design concepts and ability to use that knowledge with visualizations, using these findings and the modified RED and Glossary as a standard, we next examined students' responses to the 'Neuron Assessment' under the same conditions.

3.5.2   Students' Abilities Probed By the 'Neuron Assessment'

Four student participants Juan, Eve, Li Na and Daniel presented ideas for gathering experimental evidence using information provided by the 'Neuron Assessment' and they created diagrams to illustrate their ideas about experimental design. First, Tables 3.4-3.7 present information from interpreting diagrams in Figures 3.3-3.6 with drawings of neurons and mitochondria before and with the "Neuron Assessment'. In these tables,

we address CM and RM identified from the expert responses (Table 3.12) for neurons and each RED component before and with the assessment. Tables 3.8-3.11 compare RC before and with the 'Neuron Assessment' for concepts pertaining to neurons and mitochondria movement and then each RED component.

A.

B (after).

C (after).

D (after).
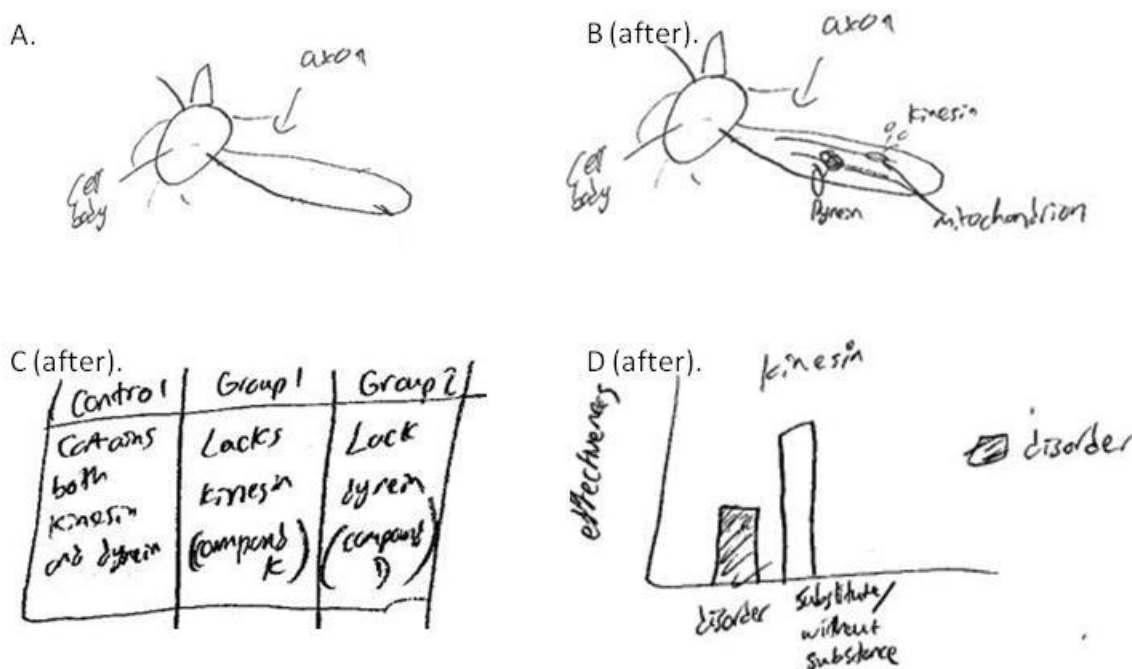
Figure 3.3 Juan's 'Neuron Assessment' figures

Before 'Neuron Assessment'.
A. Neuron concepts: Spatially manipulate an ER to interpret and explain the concept of neuron knowledge with neuron anatomy. Visualize levels of organization, relative size, shape and scale of cell body and axon. Organelle movement: Lack of Evidence (No mitochondria or organelle movement is represented). Figures depict no experimental design skills.
With 'Neuron Assessment'.
B. Neuron concepts: Decode the symbolic language composing provided 'Neuron Assessment' figures. Spatially manipulate figure of a neuron to explain knowledge of kinesin, dynein and a mitochondrion. Visualize levels of organization, relative size, shape and scale of cell body, axon, motor proteins and mitochondrion. Organelle movement in neurons: Lack of Evidence (No organelle movement is represented).
C. RED Areas: Interpret provided visuals to design experimental groups. Construct an ER to represent manipulation of control and treatment variables organized as separate groups.
D. Construct a graph (graph is flawed as inappropriate independent variables are represented on x-axis). Translate horizontally across experimental table to experimental graph (The groups represented in the experimental table do not correspond to the bars on the graph).

A (before).

O mitochondria O nucleus

cell
(neuron)   axon
moving
information

B (before).

MRI

λ → neuron
interactions

brain   micro-
scope
cell   O → mitochondria

C (after).

Control

Δ → kinesin
□ → Dynein
nucleus

mitochondria

Healthy
neuron

Experimental

Δ → kinesin
□ → Dynein
→ mitochondria
nucleus

Unhealthy
neuron

D (after).

Control	Experimental			
	Kinesin	Dynein	Kinesin + Dynein	No proteins
	Inject compound K	Inject compound D	Inject both K + D	

E (after).

Amt of movement
without added proteins

Amt of movement
compd K   compd D   compd K + D
← PROTEINS →

Figure 3.4: Eve's 'Neuron Assessment' figures

Before 'Neuron Assessment'.
A. Neuron concepts: Neuron knowledge: Spatially manipulate an ER to interpret and explain the concept of neuron knowledge with neuron cell body and axons. Visualize levels of organization, relative size and scale (relative size and shapes of cell body and axon).
Organelle movement in neurons: Spatially manipulate an ER to interpret and explain a concept. Mitochondria represented in the cell body but its movement (for example by using arrows) is not represented.
B. RED Areas: Visualize levels of organization, relative size and position of neurons relative to the organ and cellular level diagrams.
With 'Neuron Assessment'.
C. Neuron knowledge: Decode the symbolic language composing provided 'Neuron Assessment' figures. Translate horizontally across provided representations of neuron and create own visuals of a neuron. Organelle movement in neurons: Lack of Evidence (no organelle movement represented in visual representation of neurons).
D. RED Areas: Interpret an ER (provided visuals) to design experimental groups. Construct an ER (experimental table constructed to represent control and treatment variables organized as separate groups). E. Construct an ER (graphical representation) with independent variable on x-axis and dependent variable on y-axis. Different treatments are represented as separate bars. Translate horizontally across experimental table and graph representing each treatment in the table as separate bars.

Juan and Eve showed consistent difficulties reasoning with modes of representation both before and with the 'Neuron Assessment'. In contrast, Li Na and Daniel, like the expert, corrected their difficulties when prompted with the 'Neuron Assessment'. Because the findings differ for these two groups, results for Juan and Eve are presented first, followed by Li Na and Daniel. Students' RM (Tables 3.4-3.7) and RC (Table 3.8-3.11) abilities were characterized according to evidence of correct ideas (green cells), of difficulties (red cells), and for lack of evidence (orange) when information was missing for subject matter or a certain RED component. In contrast to the scientist, students in this case study provided clear evidence of their difficulties, and the degree of difficulties varied across these four students as indicated by prevalence of red cells.

3.5.2.1 Students' Reasoning with Visualizations of Experimental Design

Students' knowledge and difficulties with modes of representation were coded using concepts from the new glossary (Appendix G; underlined parts show modifications).

Table 3.4: Juan's reasoning with visualizations (RM) before and with 'Neuron Assessment'

Concepts	RM[*]	Before	With
Neuron subject matter			
a. Neuron knowledge	**Spatially manipulate a representation**	A neuron is spatially manipulated to explain neuronal anatomy (Figure 3.3A)[#].	A neuron is spatially manipulated to explain knowledge of its anatomy with kinesin, dynein and mitochondrion (Figure 3.3B).
	Visualize levels of organization	Relative size and scale of neuron cell body and axon depicted (Figure 3.3A).	Relative size and shapes of cell body, axon, motor proteins and mitochondrion depicted (Figure 3.3B).
	Decode a representation		Decode the symbolic language composing provided 'Neuron Assessment' figures (Figure 3.1a-c).
b. Organelle movement in neurons		Lack of evidence as no mitochondria or organelle movement represented (Figure 3.3A).	Lack of evidence as no organelle movement represented in visual representation of neurons (Figure 3.3B).
RED areas			
c. Experimental design table RED areas: Control group[2θ] Treatment group[18]	**Interpret a representation**	Lack of evidence	'Neuron Assessment' figures were interpreted to design experimental groups (Figure 3.1a-c).
	Construct a representation		Experimental table constructed to represent manipulation of control and treatment variable groups (Figure 3.3C).
d. Graphs RED areas: Manipulation of variable[17] Measurement of outcome[7] Accounting for variability[22] Scope of inference[15]	**Construct a representation**		Constructed graph is flawed as inappropriate independent variables are represented on x-axis [2, 3]. Bars on the graph do not correspond to the experimental table and carry no error bars [4] (Figure 3.3D).
	Translate horizontally across a representation		Experimental table translated inappropriately into a graph as the experimental table groups do not correspond to the bars on the graph[5] (Figure 3.3D).

Correct ideas   Difficulties   Lack of evidence

---

[*] RM from the list in Table 3.1

[#] Numbers and letters in parentheses refer to the Juan's diagrams in Figure 3.3

[θ]Superscripts refer to the concepts listed in column 1 and defined in Appendix G.

Before the 'Neuron Assessment', when asked about neurons and organelle movement, Juan and Eve both showed spatial manipulation in their own neuron diagrams and they visualized orders of relative scales for various anatomical parts (Figures 3.3A and 3.4A). However they struggled to represent organelle movement as Juan showed no diagrams of an organelle before being given the "Neuron Assessment' while Eve did not show any spatial manipulation as her diagrams represent mitochondria but fail to show movement (Figure 3.4A). Juan showed no evidence in his diagrams of reasoning about RED areas without the assessment (Table 3.4, row b). Eve depicted neurons in a MRI scan at the organ level (Figure 3.4A) and then zoomed in to a microscopic image (Figure 3.4B). Hence, Eve represented these visualizations across orders of magnitude (Table 3.5, row a).

Once he was given the 'Neuron Assessment' (Figure 3.1), Juan demonstrated a range of visual abilities as he decoded the provided diagrams and spatially manipulated his own images of neurons and organelle movement using appropriate orders of relative size and scale (Figure 3.3B and Table 3.4 row a). However he did not depict any organelle movement after being given the 'Neuron Assessment' (Table 3.4 row b). Similarly, Eve decoded the provided neuron diagrams (Table 3.5 row a). With the 'Neuron Assessment', she spatially manipulated her diagrams to represent anatomical parts and motor proteins kinesin and dynein with a neuron cell (Figure 3.4C) but still did not represent any movement of organelles in neurons (Table 3.5, row b). For RED areas, Juan was able to construct an experimental table (Figure 3.3C) but showed difficulties with horizontal translation from table to graph as there was a mismatch for experimental groups between the table and graph (Figure 3.3D and Table 3.4 row d). In contrast, Eve demonstrated correct RM abilities as she was able to construct an experimental table as well as designing the corresponding graph (Figure 3.4D-E and Table 3.5 row c and d).

Table 3.5: Eve's reasoning with visualizations (RM) before and with 'Neuron Assessment'

Concepts	RM[*]	Before	With
		Neuron subject matter	
a. Neuron knowledge	**Spatially manipulate a representation**	Manipulated figures of a neuron to explain knowledge of neuron anatomy (Figure 3.4A)[#].	
	Visualize levels of organization	Depicted relative size of neuron cell body and axon (Figure 3.4A).	
	Decode a representation		Decoded the symbolic language composing provided 'Neuron Assessment' figures (Figure 3.1a-c).
	Translate horizontally across representations		Translated across provided representations of neuron and created own visuals of a neuron (Figure 3.4C).
b. Organelle movement in neurons	**Spatially manipulate a representation**	Spatial manipulation is flawed as mitochondrion is depicted in cell body but shows no movement (for example by using arrows) (Figure 3.4A).	Lack of evidence as no organelle movement represented in neuron figures (Figure 3.4C).
		RED Areas	
c. Experimental design table/figure Control group[2θ] Treatment group[18]	**Visualize levels of organization**	Relative size and scale of neurons depicted at the organ and cellular level (Figure 3.4B).	
	Interpret a representation		Provided 'Neuron Assessment' figures are used to design experimental groups (Figure 3.4D).
	Construct a representation		Experimental table represents control and treatment group variables[2] (Figure 3.4D).
d. Graphs RED areas: Manipulation of variables[17] Measurement of outcome[7] Accounting for variability[22] Scope of inference[15]	**Construct a representation**	Lack of Evidence as no graph was drawn (Figure 3.4B).	Graph drawn with independent variable on x-axis and dependent variable on y-axis [2, 3.] Different treatments are represented as separate bars (Figure 3.4E).
	Translate horizontally across representations		Experimental table translated graphically with treatments shown as separate bars on the graph appropriately[5] (Figure 3.4E).

▇ Correct ideas	▇ Difficulties	▇ Lack of evidence

---

[*] RM from the list in Table 3.1

[#] Numbers and letters in parentheses refer to the Eve's diagrams in Figure 3.4 on page 102

[θ] Superscripts refer to the concepts listed in column 1 and defined in Appendix G

In the following sections, Tables 3.6 and 3.7 compare how well student abilities with modes of representation (RM abilities) in RED areas are demonstrate before being given the 'Neuron Assessment' and then with the 'Neuron Assessment'.

Li Na and Daniel. Before the 'Neuron Assessment', both Li Na and Daniel were able to demonstrate a range of RM abilities as they drew diagrams of a typical neuron with relative sizes for various anatomical parts but failed to depict any organelle movement (Figure 3.5A and 3.6A). Regarding RED areas, Li Na did not provide any visualization but Daniel constructed a representation of experimental groups in Figure 3.6B by drawing impaired and healthy patients. With the 'Neuron Assessment', both were able to decode neuron and organelle movement diagrams and translate between neuron images provided (Table 3.6 and Table 3.7 row a). They also decoded organelle movement in provided diagrams (Table 3.6 and Table 3.7 row b). In terms of RED areas, both represented correct visual skills as they were able to construct an experimental table with appropriate groups (Figure 3.5B and 3.6C; Table 3.6 and 3.7 row c). They also represented corresponding experimental findings using graphs (Figure 3.5C and 3.6D; Table 3.6 and Table 3.7 row d).

A (before).

mitochondria

outside

nerve end

transports signal

Brain nerve system

↓ signal

inside cell
membrane
outside cell

change negative
↓
positive

positive
↓
negative

B (after).

Control group
① Inhibiting kinesin + Inhibiting dynein

Experimental group
② Activating kinesin + Inhibiting dynein
③ Activating " + Activating "
④ Inhibiting " + Inhibiting "

C (after).

(retrograde −) (anterograde +) displacement

Figure 3.5: Li Na's 'Neuron Assessment' figures

Before 'Neuron Assessment'.
A. Neuron concepts: Spatially manipulate an ER to interpret and explain the concept of neuron knowledge with neuron anatomy. Visualize levels of organization, relative size and scale of nervous system, cell body, axon and mitochondria). Interpret the temporal resolution of ERs (shows signal transmission across cell as mode of neuron communication). Organelle movement in neurons: Lack of Evidence (Mitochondria are represented but movement of signals are depicted rather than of mitochondria).
Lack of Evidence (Figure shows no evidence for experimental design skills).
With 'Neuron Assessment'.
Neuron concepts: No additional diagrams drawn. Organelle movement in neurons: No additional diagrams drawn.
B. RED Areas: Interpret provided neuron visuals to design experimental groups. Construct a graph to represent manipulation of control and treatment variables organized as separate groups. Note that treatments 1 and 4 are identical. Treatment 4 was meant to be inhibiting kinesin and activating dynein.
C. Interpret provided visuals to design experimental groups. Construct a graph to represent control and treatment variables organized as separate groups; independent variables and dependent variables are represented on x-axis and y-axis respectively. Translate horizontally across experimental table and experimental graph representing each treatment in the table as separate bars on the graph appropriately.
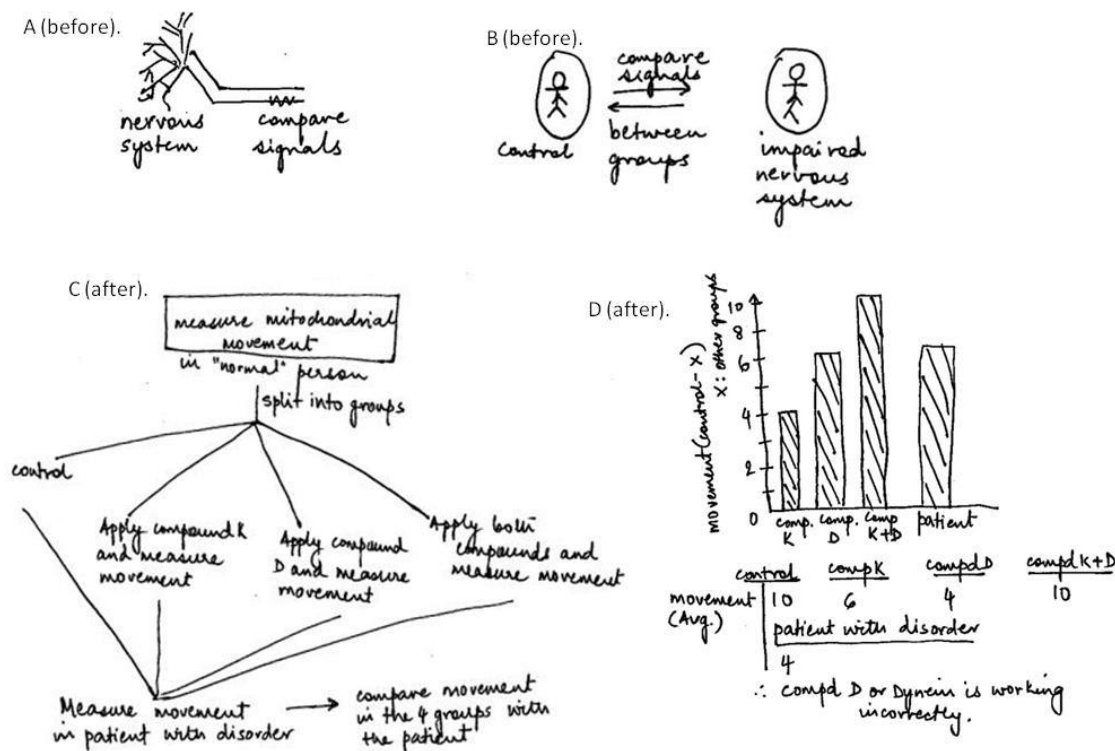
Figure 3.6: Daniel's 'Neuron Assessment' figures

Before 'Neuron Assessment'.
A. Neuron concepts: Spatially manipulate a representation to interpret and explain the concept of neuron knowledge with neuron anatomy. Visualize levels of organization, relative size and scale of axon and dendrites. Organelle movement in neurons: Lack of Evidence (No representation was created to depict organelle movement).
B. RED Areas: Construct a representation to explain experimental groups considered and measurement of outcome.
With 'Neuron Assessment'.
Neuron concepts: No new diagrams drawn. Organelle movement in neurons: No new diagrams drawn.
C. RED Areas: Interpret a representation (provided neuron visuals) to design experimental groups. Construct a representation to represent manipulation of control and treatment variables organized as separate groups.
D. Construct a graph representation with independent variables and dependent variables on x- and y-axes respectively. Different treatment groups are represented as separate bars. Translate horizontally across experimental table and experimental graph representing each treatment in the table as separate bars on the graph appropriately.

Table 3.6: Li Na's reasoning with visualizations (RM) before and with 'Neuron Assessment'

Concepts	RM[*]	Before	With
		Neuron subject matter	
a. Neuron knowledge	**Spatially manipulate a representation**	Figure of drawn neuron was used to explain knowledge of neuron anatomy (Figure 3.5A)[#].	Lack of evidence as no new representations were created to depict neurons.
	Visualize levels of organization	Relative size and shapes of cell body, axon and mitochondria depicted (Figure 3.5A).	
	Interpret the temporal resolution of a representation	Showed signal transmission as a mode of neuron communication (Figure 3.5A).	
b. Organelle movement in neurons		Lack of evidence as figures depict movement of signals but no movement of mitochondria (Fig. 3.5A).	Decoded the symbolic language composing provided 'Neuron Assessment' figures (Figure 3.1a-c).
		RED Areas	
a. Experimental design table/figure Control group[2θ] Treatment group[18]	**Interpret a representation**		Provided neuron visuals were used to design experimental groups (Figure 3.5B).
	Construct a representation		Table constructed to depicted manipulated control and treatment variable groups (Figure 3.5B).
b. Graphs RED areas: Manipulation of variables[17] Measurement of outcome[7] Accounting for variability[22] Scope of inference[15]	**Interpret a representation**	Lack of evidence	Provided 'Neuron Assessment' figures were used to design experimental groups (Figure 3.5C).
	Construct a representation		Graph constructed to represent control and treatment variable groups and independent variables and dependent variables were represented on x- and y-axes respectively)[2,3] (Figure 3.5C).
	Translate horizontally across a representation		Experimental table was translated into a graph representing each treatment in the table as separate bars appropriately (Figure 3.5C)[5].

	Correct ideas		Difficulties		Lack of evidence

---

[*] RM from the list in Table 3.1
[#] Numbers and letters in parentheses refer to the Li Na's diagrams in Figure 3.5
[θ] Superscripts refer to the concepts listed in column 1 and defined in Appendix G

Table 3.7: Daniel's reasoning with visualizations (RM) before and with 'Neuron Assessment'

Concepts	RM[*]	Before	With
Neuron subject matter			
a. Neuron knowledge	**Spatially manipulate a representation**	Figure of a neuron manipulated to explain knowledge of neuron anatomy (Figure 3.6A)[#].	Lack of evidence as no new representations were created to depict neurons.
	Visualize levels of organization	Relative size and shapes of axon and dendrites depicted.	
b. Organelle movement in neurons		Lack of Evidence as no depiction of organelle movement (Figure 3.6A).	Decoded the symbolic language composing provided 'Neuron Assessment' figures (Figure 3.1a-c).
RED areas			
c. Experimental design table/figure Control group[2θ] Treatment group[18]	**Interpret a representation**		Provided neuron visuals were used to design experimental groups (Figure 3.6C).
	Construct a representation	Experimental groups[2] considered and measurement of outcome[3] (Figure 3.6B).	To represent manipulation of control and treatment variables groups[2] (Figure 3.6C).
d. Graphs Manipulation of variables[17] Measurement of outcome[7]	**Construct a representation**	Lack of Evidence as no graph was drawn (Figure 3.6B).	Graph constructed with independent variables and dependent variables[2,3] on x- and y-axes respectively. Different treatment groups are represented as separate bars (Figure 3.6D).
	Translate horizontally across representations		Experimental table translated into graph representing each treatment in the table as separate bars appropriately (Figure 3.6D).

☐ Correct ideas   ☐ Difficulties   ☐ Lack of evidence

---

[*] RM from the list in Table 3.1
[#] Numbers and letters in parentheses refer to the Daniel's diagrams in Figure 3.6
[θ] Superscripts refer to the concepts listed in column 1 and defined in Appendix G

To summarize, before students got the 'Neuron Assessment', all four showed no evidence of depicting any movement of mitochondria along neurons and also no graphical representations of experimental results. However with the 'Neuron Assessment', Eve, Li Na and Daniel were able to interpret the supportive diagrams and create their own experimental design tables and graphs but Juan showed difficulties (Table 3.4 row d) when his 'Neuron Assessment' response revealed no evidence of mitochondrial movement and clear evidence of difficulty with constructing a graph.

3.5.2.2 Students' Reasoning with Concepts of Experimental Design

The students presented knowledge of the subject matter and experiments as they explained investigations designed to study a disorder with mitochondrial movement in neurons. Tables 3.8-3.11 show knowledge and difficulties with subject matter and experimental design before and with the 'Neuron Instrument'. We characterized correct ideas (green boxes) and difficulties (red boxes) with concepts relevant to mitochondrial movement and each component of the RED. For example, Juan's considerations for measurement of outcome *("Scientists would be measuring the degree of necessity of a certain motor protein")* showed evidence of difficulty (Table 3.8, 3.a) with concept of a variable and outcome variable (Appendix K) as "degree of necessity" is not a measurable outcome (Appendix I, Page 9, RED, Area of Difficulty 3-e). A superscript number for each concept corresponds to the glossary list (Appendix G). RC abilities in adjacent columns show what students did or how the concept was used at each stage of 3P SIT. Evidence was identified either when the students used the specific 'term' or provided an explanation that indicated knowledge of the concept as defined in the glossary.

Table 3.8: Juan's abilities with reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
		Neuron subject matter			
(I) Neuron concepts	a. Neuron knowledge	*"Neuron has an axon. And mitochondria".*	Memorize parts of neuron anatomy.	*"I am familiar with how a neuron looks with axons."*	Memorize parts of neuron anatomy.
	b. Organelle movement		Lack of evidence	*"Scientists want to see if kinesin or dynein malfunction is responsible in causing the disorder. Anterograde and retrograde movement in neurons takes place with help of kinesin and dynein".*	Apply knowledge of neurons, molecules like kinesin, dynein and mechanisms like antero- and retrograde movement to explain investigation goal of diagnosing impaired mitochondrial movement mechanism.
		RED areas			
(1) Variable property of experimental subject	a. Experimental subject *Sample[13]#* *Subject[16]* *Unit[19]* *Variable[20]*	*"[Scientists] would do individual experiments on mitochondria, kinesin and dynein[16]. They could remove kinesin and see that the mitochondria will only move[20] one way."*	Integrate knowledge of neuron concepts (mitochondria, kinesin, and dynein) [16] with the experimental subject and its variable property (movement of mitochondria) [20].	*"Neurons [16, 19] that lack kinesin [20] and neurons that lack dynein".* (RED, Area of Difficulty 1-b)	Apply knowledge of neuron concepts (kinesin and dynein) to propose a variable property of the experimental subject[16]. The variable property (neurons lacking kinesin) [20] is not aligned to the investigation goal.
(2) Manipulation of variables	a. Treatment variable *Subject[16]* *Treatment variable[17]* *Treatment group[18]*	*"[Scientists] could remove kinesin[17, 18] and see that the mitochondria[16] will only move one way."*	Integrate knowledge of experimental subject[16] (kinesin, mitochondria) to propose treatment variables[17] (removal of kinesin).	*"Use compound K [17] on neurons that lack kinesin [18] and compound D[17] on neurons that lack dynein[18]".* (RED, Area of Difficulty 2-d)	Reason globally about treatment variables[17] (treatment with compound K to neurons lacking kinesin[18] confounds the experimental goal of investigating the disorder).
	b. Control variable *Control[1]* *Control group[2]*		Lack of evidence	*"They will select a patient with a disorder as control and one without the disorder and*	Transfer and apply knowledge of concept of control groups[2] for comparison purposes.

Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.8: Juan's abilities with reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
				compare [2]". (RED, Area of Difficulty 2-j)	
	c. Controlling outside variables *Confounding variables[8]#* *Control group[2]* *Treatment group[18]*		Lack of evidence		Lack of evidence
(3) Measurement of outcome	a. Outcome *Variable[20]* *Subject[16]* *Outcome variable[7]*	*"[Scientists] would be measuring the degree of necessity of a certain motor protein[7, 20]".* (RED, Area of Difficulty 3-e)	Apply knowledge of the concept outcome variable[7, 20] to propose a suitable measure.	*"They would be measuring movement[7] of mitochondria to see if it changes without the protein".* (RED, Area of Difficulty 3-e)	Apply knowledge of the concept outcome variable[7] to propose a suitable measure. *No specific outcome proposed here (measurement of change in movement is not specific indication of a measure).*
(4) Accounting for variability	a. Replication *Variability[22]* *Subjects[16]* *Units[19]* *Treatment group[18]* *Control group[2]*				
	b. Randomization *Randomization[11]* *Random sample[10]* *Treatment groups[18]* *Variability[22]*		Lack of evidence		Lack of evidence
	c. Representative sample *Random sample*				

Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.8: Juan's abilities with reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
	[10] *Control group* [2] *Treatment group*[18]				
(5) Scope of inference	a. Scope of Inference[15#]	*"If [scientists] find a problem with kinesin and/or dynein, they could manufacture genetically some substitute for the missing motor proteins and observe the effect[15]".* (RED, Area of Difficulty 5-b; 5)	Reason locally (replacing genetically modified kinesin with impaired kinesin) and globally to make appropriate inferences[15] from experimental findings (scope of inference for patients with a neuronal disorder).		Lack of Evidence
	b. Cause and effect *Treatment Variable*[17] *Outcome variable* *Confounding Variables* [8] *Correlations*[3]		Lack of Evidence	*"When kinesin is lacking and thus, replaced with a genetically modified version of kinesin protein[17], the patient showed improvement in mitochondrial movement[7, 3,8]".* (RED, Area of Difficulty 5-c)	Apply knowledge of treatment[17], control[1] and outcome[7] variables to develop causal [3,4] explanations (causal explanations are made with respect to a mismatched treatment variable and no variability measures are considered)

▇ Correct ideas ▇ Difficulties ▇ Lack of evidence

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.9: Eve's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
		Neuron subject matter			
(I) Neuron concepts	a. Neuron knowledge	*A neuron is connected to other axons to distribute information*	Memorize knowledge of neurons and axons.	*In psychology I have seen similar types of neurons and axons in the brain.*	Apply knowledge of neurons to interpret the experimental context.
	b. Organelle movement	*What's going on in the mitochondria determines how [organelle] transport occurs".* (RED, Area of Difficulty 1-b)	Reason locally (mitochondrial process) and globally (processes inside mitochondria regulate organelle movement).	*People with the disorder are unable to perform transport and scientists believe that it has to do with motor proteins- kinesin and dynein not working and it effect on movement of mitochondria.*	Apply knowledge of concepts like transport, kinesin, dynein, mitochondria to explain the investigation goal.
		RED Areas			
(1) Variable property of experimental subject	a. Experimental subject *Sample[13#] Subject[16] Unit[19] Variable[20]*	*"[Scientists] would have to take a living specimen of the neurons[13, 16] and keep it in the environment to function properly and observe how it affects overall transport[20]."*	Apply knowledge of the neuron[16] concepts (living cells) to propose experimental subjects and its variable property[20] (transport).	*"You can try a neuron with only kinesin[16, 20] and inject compound K".* (RED, Area of Difficulty 1-b)	Apply knowledge of experimental subject[16] but the variable[20] property is not aligned with the investigation goal (impaired neurons with only kinesin with not allow unbiased investigation of whether kinesin and/or dynein are the source of the neuron disorder).

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.9: Eve's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
(2) Manipulation of variables	a. Treatment variable[17] *Treatment group[18]*	*"[Scientists] would inject[17] what they need, to manipulate things in the processes of the neurons."*	. Apply knowledge of the treatment variable[17] (injection of compounds)	*"Add compound K [17] to neurons with only kinesin[18]; compound D to neurons with only dynein".* (RED, Area of Difficulty 2-d)	Reason locally (inject compound K to neurons only carrying kinesin) and globally (using neurons with only kinesin confounds the experimental goal of investigating whether kinesin or dynein are responsible for the neuron disorder) about treatment variables [17]
	b. Control variable *Control[1#] Control group[2]*	*"[Scientists] are going to need the control[1, 2] which would be people that don't have the disorder so healthy neurons and experiment would be people that carry the unhealthy neurons."* (RED, Area of Difficulty 2-j)	Reason globally about control[1, 2] (Experimental subjects carrying obvious differences are assigned to experimental vs. control group.)	*"Neurons without any proteins[2] [kinesin or dynein]".* (RED, Area of Difficulty 2-h)	Transfer and apply knowledge of control (control group[2] does not provide natural behavior conditions because absence of the manipulated variable in treatment group, results in conditions unsuitable for the experimental

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.9: Eve's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
				subject.)
c. Controlling outside variables *Confounding variables[8]* *Control group[2]* *Treatment group[18]*		Lack of evidence	*"Neurons in control[2] and experimental group[18] with both carry same organelles[8]"*	Apply knowledge of controlling confounding variables[8] to have uniform experimental subjects in control[2] and treatment[18] groups.
(3) Measurement of outcome	a. Outcome *Variable[20]* *Subject[16]* *Outcome variable[7]*	*"[Scientists] would observe to see what happens if they specifically change a certain thing[7]"*. (RED, Area of Difficulty 3-f)	Apply knowledge of outcome variable[7] to propose a suitable measure.	*"Measure mitochondrial[16] movement[7, 20] [after treatment with compound K and D each] and compare with healthy amount of movement[7]"*. (RED, Area of Difficulty 3-e)	Apply knowledge of outcome variable[7] to propose a suitable measure. *No specific outcome proposed here (healthy amount of movement is not specific indication of a measure).*
(4) Accounting for variability	a. Replication[12] *Variability[22]* *Subjects[16]* *Units[19]* *Treatment group[18]* *Control group[2]*	*"[Scientists] have to get a significant amount of samples to test. But you need to do the experiment multiple times and so you would have to have a decent amount of neurons[16] from the healthy and unhealthy patients to conduct the experiment to compare if results are significantly close to each other[22], otherwise the*	Apply knowledge of replication[12] to propose multiple trials of the experiment but at another time as measure of dealing with variability[22].		Lack of evidence

106

Table 3.9: Eve's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
	experiment really wouldn't be accurate. Multiple trials must be done."			
b. Randomization[11#] *Random sample [10]* *Treatment groups[18]* *Variability [22]*		Lack of evidence		Lack of evidence
c. Representative sample *Random sample [10]* *Control group [2]* *Treatment group[18]*		Lack of evidence	*"The control will be the healthy neuron[2] but experimental group will be the unhealthy neurons[10, 18]".* (RED, Area of Difficulty 1-b)	Apply knowledge of representative (random) sample [10] to treatment[18] and control[2] group subjects.
a. Scope of inference[15]		Lack of evidence	*"When you see movement with kinesin and dynein inhibitor is equal to the control movement of healthy cell, your experiment is successful"[15].* (RED, Area of Difficulty 5-c)	Reason locally (presence of inhibitors) and globally (treatment with kinesin/dynein inhibitors will result in healthy neuron movements) about experimental inferences[15] don't align with provided background.
b. Cause and effect[4] *Treatment Variable [17]*	*"[Scientists] inject what they need to [17] manipulate things to see what happens if they*	Reason globally about causal claims (a causal relationship is claimed	*"With the [presence of] proteins individually, there might be loss in mitochondrial*	Reason locally (presence of inhibitors) and

(5) Scope of inference

Table 3.9: Eve's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
Outcome Variable [7] Correlations[3]	*specifically change a certain thing- and how it affects the overall transport[7]".* (RED, Area of Difficulty 5-c)	even though the data only show correlational[3] association between variables.)	*movement. But with both inhibitors [17], that is going to have full movement close to the control"[3, 4, 7,8].* (RED, Area of Difficulty 5-c)	globally (presence of inhibitors will result in healthy neuron movements) about causal relationship between treatment[17] and outcome variables[7] that do not align with provided background.

■ Correct ideas   ■ Difficulties   ■ Lack of evidence

The RC analysis revealed difficulties or lack of evidence with concepts related to both mitochondrial movement in neurons and components of the RED. In brief, for Juan and Eve, RC abilities before and with the 'Neuron Assessment' indicated that while there were some positive modifications to their knowledge, most of their difficulties before the assessment were consistent even when given the 'Neuron Assessment'. In contrast, Li Na and Daniel showed many more correct ideas when given the 'Neuron Assessment'. Concepts that showed 'lack of evidence' were developed into knowledge when they were probed with the 'Neuron Assessment'. Below is a detailed account of the interview findings from the four students. The raw transcript of these interviews has been included as Appendix K.

Juan and Eve. Both neuron subject matter and the five RED areas are shown in Table 3.8 and Table 3.9. Without the 'Neuron Assessment', both correctly depicted knowledge of a neuron (Table 3.8, I.a; Table 3.9, I.a) but showed flawed or lack of knowledge about organelle movement in neurons (Table 3.8, I.b; Table 3.9, I.b). When probed to think about how scientists discovered this information, both chose to describe experiments they may have carried out which demonstrates ability to reason with concepts of experimental design. Their descriptions provided evidence of their existing knowledge for RED areas. Both integrated knowledge of subject matter concepts to propose the variable property of experimental subject (Table 3.8, RED areas 1.a; Table 3.9 RED areas 1.a). For manipulation of variables they presented mixed responses (Table 3.8, RED areas 2a-2c; Table 3.9, RED areas 2a-2c). Both appropriately applied knowledge of the treatment variable (Table 3.8, RED areas 2.a; Table 3.9, RED areas 2.a) but Eve had difficulties with reasoning globally about control groups (Table 3.9; RED areas 2.b) while Juan showed lack of evidence for controls (Table 3.8, RED areas 2.b) . Both participants also provided no information to control confounding variables in the study they proposed (Table 3.8, RED areas 2.c; Table 3.9, RED areas 2.c). Both showed difficulties applying knowledge of an outcome variable to propose suitable measures (Table 3.8, RED areas 3.a; Table 3.9, RED areas 3.a). They shared no knowledge about ways to account for variability like replication, randomization and using a representative

sample (Table 3.8, RED areas 4a-c; Table 3.9, RED areas 4a-c). Eve presented a difficulty with failure to show replication (Table 3.9, RED areas 4.a). For Juan and Eve, flaws with knowledge of manipulation of variables and accounting for variability resulted in missing or deficient scope of inference and causal claims that didn't align with the with goal for the investigation (Table 3.8, RED areas 5.a and 5.b; Table 3.9, RED areas 5.a and 5.b)

With the 'Neuron Assessment', both Juan and Eve correctly interpreted the assessment context and supporting figures (Table 3.8, I.a; Table 3.9, I.a). When asked about how scientists would find the cause of the disorder, they suggested designing an experiment. When probed to elaborate ideas about how one would specifically plan that experiment, he had difficulty with (1) knowledge of neuron concepts (Table 3.8, I.a). Juan described experimental procedures that revealed problems in all five RED areas with reasoning about treatment variables, and knowledge of control variables (Table 3.8, RED areas 2.a; 2.b); Apply knowledge of outcome variables to propose a suitable measure. (Table 3.8, RED areas 3.a); (4) No evidence was provided to show how variability measures would be handled (Table 3.8, RED areas 4a-c); (5) No causal conclusions would be possible from Juan's experimental design owing to missing variability measures and inappropriate treatment suggestions (Table 3.8, RED areas 5.b) . Even though Eve demonstrated correct knowledge of neurons and organelle movement along neurons (Table 3.8, I; 1.b), when she designed an experiment, difficulties with concepts belonging to four RED areas became apparent (Table 3.9, RED areas 1-5, 2.a-b,3-5). But she showed correct ideas for controlling outside variables (Table 3.9, RED area 2.c). Correct knowledge was shown for variable property of the experimental subject (Table 3.9, RED areas 1.a; 2.a-c; 3.a; 4.c; 5.a-b). She also showed lack of evidence for replication and randomization (Table 3.9, RED areas 4.a-b).

In summary, before the 'Neuron Assessment', Juan's difficulties with RC abilities in all five RED components were consistent with difficulties revealed with the 'Neuron Assessment'. Without the assessment, Eve was able to reason about the experimental subject but showed difficulties with manipulation of variables, measurement of outcome,

accounting for variability, and scope of inference. With the 'Neuron Assessment', she was able to reason with knowledge of experimental subject overall and about controlling outside variables as part of accounting for variability. But Eve still revealed difficulties with at least one or more concepts under four RED areas, manipulation of variables, measurement of outcome, accounting for variability and scope of inference.

When both Juan and Eve were asked to critically evaluate their experiment with the 'Neuron Assessment' (Phase 3 of 3P SIT), both found the 'Neuron Assessment' background *easy* to decipher (*"the background does sum up the basics")*. However they asserted that designing an experiment was rather difficult when they did not know an expected outcome as was the case for the 'Neuron Assessment' when Eve said *"It is very difficult to come up with an experiment if you don't understand what you are supposed to find out eventually".*

Table 3.10: Li Na's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
		Neuron subject matter			
(I) Neuron concepts	a. Neuron knowledge	*"Neurons transfer signals [...] the neuron can transmit that information to your brain"*	Memorize knowledge of 'signal transmission' and 'neurons'	*"Neurons have different terminals like cell terminal and there is a cell body"*	Memorize knowledge of 'neuron anatomy'
	b. Organelle movement	*"Neurons communicate with each other and gradual change in ions across a membrane help in transmitting signals along axons"* (RED, Area of Difficulty 1-b; 1-c)	Apply knowledge of neuron concepts to explain organelle movement	*"Mitochondria are along the axon of a neuron. Kinesin and dynein can cause movement in different directions of mitochondria"*	Apply knowledge of neuron concepts to explain organelle movement
		RED Areas			
(1) Variable property of experimental subject	a. Experimental Subject *Sample[13#] Subject[16] Unit[19] Variable[20]*	*"[Scientists] would amplify the process [in the neuron] [16] and label some important organelles[20]"*	Integrate knowledge of neuron[16] knowledge (neuron, organelles) to propose experimental subject and its variable property[20] (amplification of neuronal process and labeling organelles)	*"The sample/subject[13, 16] is the mitochondria in the neuron and kinesin/dynein is the variable which will be either inhibited or activated[20]"*	Apply knowledge of the neuron[16] (mitochondria, neurons, kinesin/dynein) to propose an experimental subject with variable property[20] (activation/inhibition of kinesin/dynein)

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.10: Li Na's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
(2) Manipulation of variables	a. Treatment variable *Subject[16]* *Variable[20]* *Treatment variable[17]* *Treatment group[18]*	*"[Scientists] might have labeled[17] the important organelles [16]"*	Transfer and apply the knowledge of treatment variables[17] applied to a treatment group [18] of experimental subjects [16]	*"Experimental groups will be: activate kinesin[20] and inhibit dynein/ activate kinesin and dynein/ inhibit kinesin and activate dynein [17, 18]"*	Apply knowledge of treatment variable[17] (kinesin and dynein inhibitors) to propose suitable treatments (activation/inhibition) applied to experimental subjects [16]
	b. Control variable *Control[1#]* *Control group[2]*		Lack of evidence	*"Neurons treated with kinesin and dynein inhibitors will be the control group [1, 2]".* (RED, Area of Difficulty 2-i)	Reason globally about control group[2] (control group needs to carry neurons in natural condition as inhibition of organelle movement in neurons will not allow comparison to treatment groups).
	c. Controlling outside variables *Confounding variables[8]* *Control group[2]* *Treatment group[18]* *Variation[21]*		Lack of evidence	*"Before the treatments subjects should have the same conditions [8, 21] in the treatment and control groups [2, 18]. Otherwise, they may react differently leading to false causation"*	Apply knowledge of the controlling outside variables8 (experimental subjects subjected to same conditions) in treatment[18] and control groups[2] as

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.10: Li Na's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
					a measure to reduce variation[21]
(3) Measurement of outcome	a. Outcome *Variable[20] Subject[16] Outcome variable[7]*	*"[Scientists will] measure which organelle will cause movement in different directions[7]; They could measure the direction and displacement or electrical potential[7, 20]"*	Apply knowledge of a specific measureable outcome[7] that the experimental subject[16] carries in response to experimental conditions (The outcome proposed here is not in response to experimental but natural conditions).	*"Displacement of mitochondria[7, 16, 20] can be measured in the form of length in micrometers"*	Apply knowledge of a specific measureable outcome[7] that the experimental subject[16] carries in response to experimental conditions
(4) Accounting for variability	a. Replication[12] *Variability[22] Subjects[16] Units[19] Treatment group[18] Control group[2]*		Lack of evidence	*"We need to use a large number of samples[16] in treatment[18] and control groups[2], to observe data outliers[22] and then just decide values that lie centrally"*	Apply knowledge of replication[12] to experimental subjects[16] (large number of samples) as measure to reduce variability[22]
	b. Randomization[11] *Random sample[10] Treatment groups[18] Variability[22#]*		Lack of evidence	*"Neurons need to be picked at random and assigned to treatments completely randomly[11, 22]. You consider that all cells are the same and randomly assign[11] them to the experimental groups"*	Apply knowledge of random sampling[10] and randomization[11] (random assignment of treatments in treatment groups[18]) as measure to reduce variability[22]
	c. Representative		Lack of evidence	*"[For both treatment[18]*	Apply knowledge

Superscripts refer to the concepts listed in column 2 and defined in Appendix G

114

Table 3.10: Li Na's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
sample[10] *Sample [13]* *Random sample [10]* *Control group [2]* *Treatment group[18]*			*and control groups[2]] I will keep the same organelles under observation[13], use the same species of organisms for the neurons and use cells from the same one animal. And also make sure that they are in the same environment"*	of selecting a representative random sample[10, 13] in the treatment[18] or control[2] group (organism species and cells) as a measure to average out variations
a. Scope of inference[15]				Lack of evidence
b. Cause and effect[4] *Treatment Variable [17]* *Outcome variable [7]* *Correlations[3]*	Lack of evidence		*When kinesin is activated and dynein is inhibited[17], we see movement in the anterograde direction[7]. When dynein is working and kinesin is inhibited[17] we see movement in the retrograde direction[7]. When both are activated, the functions of the two proteins are replicated and thus, the mitochondria cannot move in either direction so the movement is impaired [3]."* (RED, Area of Difficulty 5-c)	Reason globally about causal claims (contradictory correlation[3] relationship between treatment[17] and outcome[7] variables is suggested)
(5) Scope of inference				

Correct ideas Difficulties Lack of Evidence

Table 3.11: Daniel's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
		Neuron subject matter			
(I) Neuron concepts	a. Neuron knowledge	i. *"Nerves carry signals throughout your body to move or other processes".*	Memorize entities: nerves and signal transmission processes	ii. *"Neurons have axons and a branched structure".*	Memorize entities: axon structure
	b. Organelle movement	i. *"I just think of electrical signals that would move against the wall of the neuron".* (RED, Area of Difficulty 1-b)	Apply knowledge of neuron concepts (signal transmission) to explain organelle movement	ii. *"Two proteins help in the movement. One protein goes one way and the other goes the other way. They move along an axon of a neuron."*	Integrate knowledge of structure and function of neuron concepts (two proteins, axon) to explain organelle movement mechanism
		RED Areas			
(1) Variable property of experimental subject	a. Experimental subject[#] *Sample[13] Subject[16] Unit[19] Variable[20]*	i. *"An experiment involving people[16] with impaired nerves[20]".*	Apply knowledge of variable[20] property (impairment of nerves) to experimental subject[16].	ii. *"There are two different compounds to inhibit two different proteins and observe which inhibited protein affects mitochondrial movement in neurons [16, 20]".*	Apply knowledge of experimental subject[16] (neurons) and variable property[20] (mitochondrial movement under the effect of proteins)
(2) Manipulation of variables	a. Treatment variable *Subject[16] Variable[20] Treatment variable[17] Treatment group[18]*	i. *"[Scientists] would compare signals[20] among people in the control groups with the experimental group[18] that have an impaired nervous system [17, 20]".*	Apply knowledge of treatment group[18] of experimental subjects[16] exposed to experimental conditions that vary[20] (varying signals in control vs. experimental groups) in a certain way.	ii. *"Split cells of normal persons into 5 different groups[18]. Each group carries a different treatment [normal person; control with no treatment, one with compound K[20] and another one with compound D[20]; one gets both]"*	Apply knowledge of treatment group[18] of experimental subjects[16] exposed to experimental conditions that vary[20] (varying compound treatments) in a certain way.

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.11: Daniel's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
	b. Control variable *Control[1]* *Control group[2]*	i. *"Comparing with a control group with people that have normal/regular nervous system[1,2]".* (RED, Area of Difficulty 2-j)	Transfer and apply the knowledge of the concept of control[1,2]. Reason globally about the concept of control[1,2] (Experimental subjects carrying obvious differences are assigned to experimental vs. control group).	ii. *"The control group[1,2] would not be receiving any treatment but would still be subjected to the same conditions as the treatment group".*	Reason globally about the concept of control[1,2] (Parameters other than the treatment variable are identical for both treatment and control conditions).
	c. Controlling outside variables *Confounding variables[8#]* *Control group[2]* *Treatment group[18]* *Variation[21]*	i. *"[Scientists] would try to keep people as similar[8,21] as possible so it's just the nervous system that's different between the two (treatment[18] and control[2]) groups so results aren't affected".*	Apply knowledge of controlling outside variables[8,21] by matching control[2] and treatment[18] groups as closely as possible.	ii. *"People (in treatment[18] and control[2] groups) need as similar as possible, in health conditions, so that we know that the observed effect is due to compound K or D application[18]".*	Apply knowledge of controlling outside variables[8,21] by matching control[2] and treatment[18] groups as closely as possible to draw clear causal claims.
(3) Measurement of outcome	a. Outcome *Variable[20]* *Subject[16]* *Outcome variable[7]*	i. *"You could measure the strength of the electrical signals or the path the signal takes[7,20]".* (RED, Area of Difficulty 3-c)	Apply knowledge of outcome variable[7] to propose a suitable measure (association of measuring strength of electrical signals with measurement of organelle movement is not explained).	ii. *"I predict with treatment of compound K, the mitochondria moved 4 units less than the control groups it over a specific period of time[7,20]".*	Apply knowledge of outcome variable to propose measureable outcomes.
(4) Accounting	a. Replication[12]	i. *"Scientists would try to*	Apply knowledge of	ii. *"I would use groups[12]*	Apply knowledge of

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.11: Daniel's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

	Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
for variability	*Variability[22]* *Subjects [16]* *Units [19]* *Treatment group [18]* *Control group[2]*	*measure the electrical signals in the two different groups [2, 12,16,18,19.]*	replicating[12] measurements in groups of experimental subjects[16] across treatment[18] and control groups[2] as a measure to reduce variability[22].	*of neurons[16] for each experimental group[2, 18 ".]*	replicating[12] measurements in groups of experimental subjects (neurons) [16] in each experimental group[2, 18] as a measure to reduce variability[22].
	b. Randomization[11#] *Random sample [10] Treatment groups [18] Variability [22]*	*i. Lack of Evidence*		ii. *"I would randomly assign cells[11] into groups[18] to avoid biasing[22] the results and only measure effect of the compounds".*	Apply knowledge of 'randomization'[11] of treatment group[18] conditions as a measure to reduce variability[22] and bias in the experiment.
	c. Representative Sample *Sample [13], Random sample [10], control group [2], treatment group[18]*	*i. Lack of Evidence*		ii. *"Use a sample of patients with the same age range, height etc[10, 13] so that only the neurons are different between the two groups[2, 18] to avoid biasing the results".*	Apply knowledge of 'representative sample' [10, 13] selection in treatment[18] and contro[12] groups as a measure to reduce bias experimental results.
(5) Scope of inference	a. Scope of Inference *Scope of Inference[15]*	i. *"If there is a difference between heights of subjects in two different groups, you wouldn't be able to necessarily decide if it was the height that gave rise to the difference in strength of the electrical signals*	Reason globally about inference[15] of experimental results (difference in electrical signal strength is an irrelevant variable and thus inferences are made to an	ii. *"Compare the movement with multiple patients who have the disorder with the 4 groups of patient. This will allow us to infer that those were the protein that caused the disorder[15]".*	Reason locally and globally (variability measures, suitable control and experimental groups, movement as the variable property and measureable outcome variable) to draw inferences[15] about the protein impairment leading to the

---

Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Table 3.11: Daniel's abilities reasoning with concepts (RC) before and with 'Neuron Assessment'

Concepts	Before (Phase 1)	RC	With (Phase 2)	RC
	rather than the nervous system[15]". (RED, Area of Difficulty 5-b)	irrelevant target population).		neuronal disorder.
b. Cause and effect[4] *Treatment Variable [17], control[1] and outcome variable[7], Correlations[3]*	i. *"You could measure the strength of the electrical signals[7] or the path the signal takes and see differences in sending signals [3,7]"*. (RED, Area of Difficulty 5-b)	Integrate knowledge of relevant measurable outcome variables[7] to draw appropriate causal explanations. Reason globally to claim a causal relationship[4#] separate from correlations[3] (measurement of electrical signals is mismatched with investigation goal).	ii. *"Compare your treatment groups"[1, 17] movements with movement in neurons of a patient with disorder to see similarities in trends of the movement. If they did have the same movement[7], you could argue the source of the disorder as per your treatment [3, 4]"*.	Reason locally (comparison of trends in mitochondrial movement in neurons) and globally (comparison of movement trends, along with variability measures lead to the protein source that leads to the neuron disorder) about the causal relationship[4] as separate from correlations[3] between treatment[17] and outcome variables[7].

■ Correct ideas ■ Difficulties ■ Lack of evidence

[#] Superscripts refer to the concepts listed in column 2 and defined in Appendix G

Li Na and Daniel. In general Li Na and Daniel performed better than Juan and Eve both before and with the 'Neuron Assessment'. Before the 'Neuron Assessment', both Li Na and Daniel accurately presented knowledge of neurons (Table 3.10, I.a; Table 3.11, I) but showed difficulty applying knowledge of organelle movement in neurons (Table 3.10, I; Table 3.11, I). Both were able to reason about experiments with concepts relevant to variable property of experimental subject (Table 3.10, RED areas 1.a; Table 3.11, RED areas 1.a), but they presented mixed abilities with knowledge of manipulation of variables. Li Na did not show any knowledge about treatment variables (Table 3.10, RED areas 2.a) in contrast to Daniel (Table 3.11, RED areas 2.a). Li Na showed no knowledge while Daniel showed difficulty applying his knowledge and reasoning about control of variables (Table 3.10, RED areas 2.b; Table 3.11, RED areas 2.b). Li Na also showed lack of knowledge about confounding variables but Daniel presented correct knowledge of this concept (Table 3.10, RED areas 2.c; Table 3.11, RED areas 2.c). Li Na presented knowledge of outcome variables with flawed outcome measures by suggesting 'displacement of mitochondria' as a measure and Daniel also had difficulty measuring dependent variables by suggesting signal strength or pathway as a measure (Table 3.10, RED areas 3.a; Table 3.11, RED areas 3.a). Li Na did not address how to deal with or measure variability (Table 3.10, RED areas 4a-c). In contrast, Daniel showed that he knew there was a need to replicate measures (Table 3.11, RED areas 4.a). Li Na did not provide evidence for reasoning about causal claims owing to lack of evidence for reporting variability in measures (Table 3.10, RED areas 5.a-b). Daniel showed difficulty with reasoning about inferences and causal claims from his experimental findings because he didn't identify appropriate measurable outcomes or proposed ways to measure variability as part of experimental findings (Table 3.11, RED areas 5.a-b).

With the 'Neuron Assessment' (Phase 2), Li Na and Daniel accurately presented their knowledge of neurons (Table 3.10, I.a-b; Table 3.11, I.a-b). Li Na also appropriately applied knowledge of RED components, variable property of

experimental subject using (Table 3.10, RED areas 1.a), measurement of outcome (Table 3.10, RED areas 3.a) and variability (Table 3.10, RED areas 4.a-c). She showed difficulty with concepts based on manipulation of variables as she struggled to reason globally about controls (Table 3.10, RED areas 2.b) and causal explanations (Table 3.10, RED areas 5.b). In contrast, Daniel sufficiently applied his knowledge of concepts from all five RED areas (Table 3.11). He also reasoned locally and globally about concepts like variability measures (Table 3.11, RED areas 4.a-c) and causal claims (Table 3.11, RED areas 5.a-b) to draw appropriate inferences from findings of his experiment after he was given the 'Neuron Assessment'.

In summary, without the assessment, Li Na showed knowledge of RED components variable property of experimental subject, measurement of outcome and accounting for variability which is also consistent with her response when given the assessment but the assessment elicited a difficulty with 'control' where there was a lack of evidence before she was given the 'Neuron Assessment'. For Daniel, without the 'Neuron Assessment' he exposed difficulties with concepts for manipulation of variables, measurement of outcome and scope of inference. Daniel corrected these difficulties when he reasoned about concepts of experimental design given the probing questions as well as the 'Neuron Assessment' background information.

As feedback (Phase 3), Li Na and Daniel both found the experimental design activity to be quite enjoyable *("I can come up with a lot of ideas so I am comfortable with activities like this")*. They also considered the background information quite useful to design an experiment *("The diagrams definitely helped me think about the process more clearly since I did not know about this process too much before this study. I think it helped me see how things like the mitochondria, kinesin, and dynein are placed within a neuron")*. Nevertheless, they expressed discomfort being uncertain if they had correctly given the expected answer for the experiment *("I don't know the right answer to this experiment so whether the question is good depends on the answer")*.

3.6 <u>Discussion</u>

In this section, patterns for expert and student modes of representations (RM) will be presented (Table 3.12) followed by patterns for reasoning with experimental design concepts (RC) (Table 3.13). Evidence suggests that the 'Neuron Assessment' is useful especially as a probe for some specific details of the RED areas.

Table 3.12: Expert and Student Reasoning with Visualizations (RM) of Experimental Design

RM[*]	Expert		Juan		Eve		Li Na		Daniel	
	Before	With	Before	With	Before	With	Before	With	Before	With
1. Decode symbolic language		x		x		x		x		x
2. Interpret and use a representation	x	x		x		x		x		x
3. Construct a representation	x	x	x	x	x	x	x	x	x	x
4. Translate horizontally among alternative representations of the same phenomenon	x	x		x	x	x		x		x
5. Visualize levels of organization	x		x	x	x	x	x		x	
6. Interpret the temporal resolution	x	x					x			
7. Spatially manipulate a representation	x		x	x	x	x	x		x	

■ Before and with 'Neuron Assessment'   ■ Before 'Neuron Assessment' only   ■ With 'Neuron Assessment' only

_____

[*] RM from the list in Table 3.1

In answer to research question 1, how well does the 'Neuron Assessment' reveal the nature of expert knowledge about organelle movement in neurons and the experiments used to elucidate that knowledge, we find that the 'Neuron Assessment' is a good probe to distinguish expert reasoning about experiments from the performance of a typical undergraduate student. In answer to research question 2, how well does the 'Neuron Assessment' expose student knowledge and related difficulties with experiments to investigate organelle movement in neurons, findings show that the assessment provided students with adequate information to demonstrate how they reason with visual representations (RM) and experimental design concepts (RC) to support their ideas about investigating a current research problem. In general findings show that Juan and Eve were typical students and did better with the 'Neuron Assessment'. Li Na, Daniel and the Scientist showed more knowledge before the 'Neuron Assessment'.

3.6.1 Expert and Student Reasoning with Visualizations (RM) of Experimental Design

Findings with nature of expert knowledge (RQ1) indicate that "spatial manipulation across representations" (Table 3.12, row 7) for experimental design could be assessed using a different sort of experiment. The MACH model development study (Trujillo et al., *in press*) showed that neurobiologist and cancer biologist infer a mechanism from experimental/temporal data whereas the structural biologist infers a mechanism from spatial research findings. In reality, all mechanisms involve both spatial and temporal changes. Yet, the current findings indicate that experimental design by the expert scientist was often interpreted without referring back to the spatial (in most cases) or temporal (in some cases) features of the neuron.

The 'Neuron Assessment' figures were suitable for expert and all students to decode the information presented (Table 3.12, row 1). All participants used information provided to construct their own figures relevant to investigations they

designed for the 'Neuron Assessment' (row 3). The assessment was good to show interpretation and use of representation and horizontal translation across representations (row 2 and 4) because three out of four students who did not show these abilities were able to do so with the 'Neuron Assessment' (RQ2).

In contrast, the 'Neuron Assessment' may not be good to show visualization of the levels of organization (Table 3.12, row 5) because before the assessment two students and the expert who visualized more about neuron anatomy and mechanisms with neurons, like organelle movement and signal transduction. However, with the assessment they continued to refer to the ideas that they had already explained. The assessment did not probe students to interpret temporal resolution as only the expert but no students did so with the assessment. In fact, with the assessment, all students chose to represent comparison groups rather than time course graphs. This indicates that the 'Neuron Assessment' is good to probe use of comparison groups and perhaps, temporal resolution may be replaced by with/without experimental comparison (control/treatment) groups.

In summary, the 'Neuron Assessment' provides useful evidence for RM abilities as the more proficient students Li Na and Daniel demonstrated visual abilities like the expert before and with the assessment. The typical students, Juan and Eve, who did not show certain visual abilities before the assessment were able to do so once they were exposed to the assessment.

Table 3.13: Expert and Student Reasoning with Concepts (RC) of Experimental Design

RC Concepts	Expert Before	Expert With	Juan Before	Juan With	Eve Before	Eve With	Li Na Before	Li Na With	Daniel Before	Daniel With
1. Neuron	x	x	x	x	x	x	x	x	x	x
2. Organelle movement	x	x		x	x	x	x (diff)	x	x (diff)	x
3. Experimental Subject	x	x	x	x	x	x (diff)	x	x	x	x
4. Variable	x	x	x	x	x	x (diff)	x	x	x	x
5. Treatment variable		x	x	x (diff)	x (diff)	x (diff)	x	x	x	x
6. Treatment group		x	x	x (diff)	x (diff)	x (diff)	x	x	x	x
7. Control variable	x	x			x (diff)	x (diff)		x (diff)	x	x
8. Control	x	x		x	x (diff)	x (diff)		x (diff)	x (diff)	x
9. Control group	x	x		x	x (diff)	x (diff)		x (diff)	x (diff)	x
10. Controlling outside variables	x	x				x		x	x	x
11. Confounding variables	x	x				x		x	x	x
12. Variation	x	x			x (diff)			x	x	x
13. Outcome variable	x	x	x	x	x (diff)	x (diff)	x	x	x (diff)	x
14. Replication	x	x			x (diff)			x	x	x
15. Variability	x	x			x (diff)			x	x	x
16. Randomization	x	x						x		x
17. Representative sample	x	x				x (diff)		x		x
18. Scope of Inference	x	x	x			x (diff)			x (diff)	x
19. Cause and effect	x	x		x (diff)		x (diff)			x	x
20. Correlations	x	x		x (diff)		x (diff)		x (diff)	x (diff)	x

■ Before and with 'Neuron Assessment'   ■ Before 'Neuron Assessment' only   ■ With 'Neuron Assessment' only

3.6.2 Expert and Student Reasoning with Concepts (RC) of Experimental Design

The context of the 'Neuron Assessment' is a good probe for all concepts in the glossary (Appendix G). Expert used those experimental design concepts to present knowledge for the 'Neuron Assessment' (Table 3.13) (RQ1). The expert revealed knowledge of treatment variables with the 'Neuron Assessment' even though this information was missing before the assessment.

The 'Neuron Assessment' is a good probe for knowledge of several experimental design concepts for students (RQ2). The assessment was good for Daniel as he showed knowledge of all concepts (Table 3.13). The assessment was poorest for concepts 12-16 (variation, outcome variable, replication, variability, randomization and scope of inference), weaker for concepts 7 (control variable), 10-11(controlling outside variables, confounding variables), 17 (representative sample) and 19 (cause and effect). The 'Neuron Assessment' is great to probe for concepts 1-6 (neuron, organelle movement, experimental subject, variable, treatment variable, treatment group), 8-9 (control and control group) and 20 (correlations).

All students presented knowledge about 'neuron', 'organelle movement', 'experimental subject' and 'treatment variables' (Table 3.13, row 1-6) before and with the assessment. The assessment also revealed knowledge of 'controls' for both low and high performing students (Juan and Li Na) as this information was not presented before the assessment (row7-9). Three of four students showed lack of evidence for several 'variability' related concepts (row 11-17) before but revealed knowledge and difficulties when given the 'Neuron Assessment'. Thus, findings indicate that while a high performing students like Daniel showed consistent knowledge before and after assessment, low performing students needed this prompt to reveal knowledge and in certain cases, difficulties with variability concepts.

The 'Neuron Assessment' is good for probing knowledge of causal outcome related concepts (Table 3.13, row 18-20). Students with lack of evidence for these concepts before the assessment revealed difficulties with the 'Neuron Assessment'.

However, Daniel showed difficulties before but was able to present appropriate knowledge comparable to the expert with the 'Neuron Assessment'.

The assessment was not so useful for exposing knowledge of a low performing student like Eve, in terms of certain variability related concepts (Table 3.13, row 12, 14-15). She showed difficult prior to the assessment and exposure to the assessment was not successful to reveal any knowledge or difficulty. Eve's findings show lack of evidence (Table 3.13, blank) which may reflect difficulty but Li Na's findings show that the 'Neuron Assessment' prompts correct knowledge as well as difficulties.

3.7 Summary

The 'Neuron Assessment' is a good assessment for exposing knowledge of abilities to call on modes of representation and concepts related to 'control' (Picone et al., 2007, Shi et al., 2010), 'variability measures' (Kanari & Millar, 2004; Kuhn & Dean, 2005) and 'causal outcomes'(Klahr, Fay & Dunbar, 1993; Schauble, 1996). The assessment yielded information about major experimental design areas outlined by our own and other previous research (Dasgupta et al., 2014, Deane et al., 2014) and also revealed visual modes of presenting these areas which contributed for modifications to our existing glossary list and the RED.

We find students with either weak or strong knowledge of experimental design abilities were uncomfortable with not knowing the right answer for the 'Neuron Assessment'. Perhaps, we should be doing a better job giving students practice with uncertain ideas that they can learn to test. This crucial aspect of training in scientific research is to develop an ability to pose testable questions and think about different ways to experimentally test these. So assessment like this should not have one answer, but rather could be useful for discussion since some of the experiments would be better capable of revealing new knowledge than others.

The 'Neuron Assessment' can be used to examine students' experimental design knowledge about a current research scenario. The assessment is particularly useful as it

levels for differences in prior knowledge by providing required information and visuals. The case study method described can be used to compare expert experimental design abilities to those demonstrated by a range of student in a first year undergraduate biology class. The CRM method of examining student responses helps go deep into the source of student difficulties to understand if they struggle with reasoning with visualization of experimental design and/or with knowledge of experimental concepts.

<u>References</u>

American Association for the Advancement of Science (AAAS) (2011) *Vision and Change in Undergraduate Biology Education*. AAAS, Washington, DC. http://visionandchange.org/finalreport

Dasgupta AP, Anderson T, Pelaez N (2014). Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties. CBE Life Sci Educ *13,* 265-284

Deane T, Nomme K, Jeffery E, Pollack C, Birol G (2014). Development of the Biological Experimental Design Concept Inventory (BEDCI). CBE Life Sci Educ *13,*540-551

Harker AR (2009). Full application of the scientific method in an undergraduate teaching laboratory: A reality-based approach to experiential student-directed instruction. J Coll Sci Teach *29*, 97-100.

Hiebert SM (2007). Teaching simple experimental design to undergraduates: do your students understand the basics? Adv Physiol Educ *31*, 82-92.

Kanari Z, Millar R (2004). Reasoning from data: How students collect and interpret data in science investigations. J Res Sci Teach *41*, 748-769.

Klahr D, Fay AL, Dunbar K (1993). Heuristics for scientific experimentation: A developmental study. Cognitive Psychol *25*, 111-146.

Knott AB, Perkins G, Schwarzenbacher R, Bossy-Wetzel E (2008). Mitochondrial fragmentation in neurodegeneration. Nature Reviews Neuroscience *9*, 505-518.

Kuhn D, Dean D (2005). Is developing scientific thinking all about learning to control variables? Psychol Sci *16*, 866-870.

Trujillo CM, Anderson TR, Pelaez N. A Model of How Different Experts Explain Biological Mechanisms. *In press.*

Picone C, Rhode J, Hyatt L, Parshall T (2007). Assessing gains in undergraduate students' abilities to analyze graphical data. Teach Issues Exp Ecol 5 (July 2007): Research #1. http://www.esa.org/tiee/vol/v5/research/picone/article.html (accessed December 20, 2013)

Salangam J (2007). The impact of a prelaboratory discussion on non-biology majors abilities to plan scientific inquiry. California State University, Fullerton. ProQuest Dissertations and Theses.

Schauble L (1996). The development of scientific reasoning in knowledge-rich contexts. Dev Psycho *32*, 102–119.

Schönborn KJ, Anderson TR (2009) A Model of Factors Determining Students' Ability to Interpret External Representations in Biochemistry', International Journal of Science Education *31*, 193-232.

Shi J, Power JM, Klymkowsky MW (2011). Revealing student thinking about experimental design and the roles of control experiments. Int J Sch Teach Learn *5*, 1-16.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. Cog Sci *19*, 97-140.

Yin, R. K. (1984). Case study research: Design and methods. Newbury Park, CA: Sage.

CHAPTER 4: VALIDATION OF THE 'NEURON ASSESSMENT' IN COMPARISON TO OTHER MEASURES OF BIOLOGY STUDENTS' UNDERSTANDING OF EXPERIMENTAL DESIGN

4.1 Abstract

To use an assessment for diagnostic purposes, it is important to validate the assessment for knowledge areas it aims to measure. The 'Neuron Assessment' is based on the current research context of a disorder associated with mitochondrial movement in a neuron. The assessment levels prior knowledge differences by providing sufficient knowledge of the context. To validate the 'Neuron Assessment' with published measures of experimental design that were used to develop the RED (Dasgupta et al., 2014), all three assessments carry prompts to create representations. Four biology undergraduate students and an expert neurobiologist provide responses to the three assessments in paper-pencil format in a single session. Responses are examined using a modified Rubric for Experimental Design (Dasgupta et al., 2014) that diagnoses visual abilities for each part of an experiment. Findings indicate that the 'Neuron Assessment' is comparable with the other two assessments as knowledge or difficulties detected across three assessments are consistent for majority of RED areas in case of each student. However, very few RED areas show variable knowledge across three assessments. Findings imply that an assessment with background story and appropriate visuals, like the 'Neuron Assessment' provides domain general skills that student may not yet have developed and is comparable with published assessments of experimental design.

4.2 Introduction

The Rubric for Experimental Design (RED) identifies five major areas of difficulties biology undergraduate students faced when designing investigations (Dasgupta et al., 2014). The 'Neuron Assessment' is based on the current research context of a disorder associated with mitochondrial movement in a neuron. The assessment levels prior knowledge differences by providing sufficient knowledge of the context so that the assessment can focus on measuring domain-general knowledge of experimental design. The assessment also provides opportunities to interpret and create representations such as graphs and diagrams. The goal of this study is to validate the 'Neuron Assessment' by comparing what is measures against published 'Shrimp' and 'Drug' assessments that were used to develop the RED. However, the published assessments did not provide any evidence of abilities to reason with representations Thus, for this study students were given an opportunity to create representations for all three assessments in order to make the assessments comparable. The modified RED, that includes visual abilities for parts of an experiment, was used to diagnose knowledge presented across the three assessments.

4.3 Background

The usefulness of an assessment probe requires validation against other assessments that have been shown as good measures of the same factors this assessment proposes to measure. In this study, an objective was to validate the 'Neuron Assessment' as a measure of experimental design knowledge characterized in the RED. The goal was to validate the 'Neuron Assessment' against two published assessments that were used to characterize student knowledge in developing the RED. If the 'Neuron Assessment' is a valid measure of domain general skills that are assessed by the RED with the 'Shrimp' and 'Drug' assessments, then students who do well on the 'Shrimp and 'Drug' assessments, will also perform well on the 'Neuron Assessment'. On the other hand, students those show difficulties with 'Shrimp' and 'Drug' assessment, will display the same difficulties in response to the 'Neuron Assessment'.

Background and diagrams provide the necessary content knowledge with the 'Neuron Assessment' and thus, we expect student performance to be same across all three assessments. In this chapter, students are prompted to show the modes of representation they use when reasoning about experiments across all three assessments. This provides an opportunity to use the provided figures will help us find out more the source of difficulties across all three assessments. If performance differs between the 'Neuron Assessment' compared with the 'Drug' and 'Shrimp' assessments, it could be that their ability to visualize the situation or their domain-specific knowledge interferes with their ability to transfer their experimentation knowledge from one context to another.

Previous research has reported difficulties with transfer of knowledge to new domains or contexts. Transfer refers to accurate application of reasoning skills acquired or expressed in one scientific context to other related scientific contexts. Studies demonstrate the inability of participants to recognize the analogous relations between two contexts unless the analogy is explicitly pointed out to them (Detterman & Sternberg, 1993). Furthermore, the ability to identify analogy with underlying concepts is greatly affected by learners' familiarity with an area. Familiarity affects whether people think deeply and identify underlying principles, or simply get caught in the surface features of the problem. On the other hand familiarity may lead people astray.

Experts have domain-specific knowledge that is content rich and deep. In contrast, domain-general knowledge is not as context dependent, and so is more easily transferable across different contexts than is domain-specific knowledge (Feltovich, Prietula, and Ericsson, 2006). Experts recognize the underlying principles and concepts of their domain-specific knowledge and thus can extrapolate ideas in a domain-general manner owing to their long standing experience in a certain domain (Chi, Feltovich, and Glaser, 1981; Ericsson, 2006). But students have trouble promoting transfer as they tend to categorize concepts into either domain-specific knowledge or consider it 'broadly applicable' (Detterman and Sternberg, 1993). The 'Neuron Assessment' attempts to measure domain general knowledge by providing domain specific background knowledge and figures about the context (Barnett and Ceci, 2002, Chen and Klahr, 1999,

Zimmerman 2000, 2007). If the reasoning tested is domain-general (Zimmerman, 2007, p.175), then the outcomes measured across the three assessments should be similar if the student is able to apply reasoning about experiments across three context areas.

Informed by the literature on the issue of transfer, the purpose of this study was to examine if abilities to reason with concepts and representations of experiments are transferred across the context of three experimental design based assessments. Thus, an effort was to validate the 'Neuron Assessment' with comparison of diagrams and concepts reported here (Appendix H) with the 'Shrimp' and 'Drug' assessment (Appendix A and B). Comparison across three assessments will also allow us to see if domain knowledge about experiments in a certain assessment is translated to other assessments in a domain general manner. The research question (RQ) we ask is as follows:

How well does students' performance on the 'Neuron Assessment' compare with their knowledge and difficulties revealed by other assessments of experimental design knowledge in biology?

To validate the 'Neuron Assessment', we evaluated if student knowledge and difficulties with RED (Rubric of Experimental Design) areas diagnosed by this assessment are comparable to those revealed by published assessments (Dasgupta et al., 2014). Specifically, the 'Neuron Assessment' was compared with two other published assessments ('Shrimp' and 'Drug Assessment') in terms of its effectiveness in probing for RED areas of difficulty using written responses by expert and student participants (RQ).

4.4 Method

Four student participants and an expert were recruited to complete three assessments ('Shrimp', 'Drug' and 'Neuron'; Appendix L) in paper-pencil format. Participants were given the option to withdraw from the study, or to leave answers blank if desired. Thus, our knowledge is restricted to what participants chose to write. All three

assessments were completed by each participant individually within an hour. The student participants are identified with pseudonyms Juan, Eve, Li Na and Daniel for confidentiality. The expert is referred to as Eric. Juan is a male Hispanic who is a chemistry major. Eve is a white female and microbiology major. Li Na is an Asian female and cell molecular biology major. Daniel is a white male and engineering major. The expert is a white, male, neurobiology research scientist.

The assessments were given to the participants in no specific sequence, although we ensured that the first assessment was the published 'Shrimp' or 'Drug Assessment' to help participants understand the task (See raw written assessment transcripts under Supplemental Material Appendix L). Data files were stored on a secured computer, and files were transferred using a secure, password protected file transfer system as per IRB protocol #1008009581.

4.4.1 Experimental Design Assessments

We used two published assessments as measures of five RED areas to compare findings with the 'Neuron Assessment' for this study. The 'Neuron Assessment' prompts students to design an experiment to investigate a disorder related to organelle movement in neurons (See Chapter 3, Figure 3.1). This assessment provides content knowledge and figures about neurons and a neurological disease. Students then apply their knowledge of RED areas such as proposing a hypothesis, considering variables to manipulate, organizing comparison groups, and reporting causal conclusions from an experiment they are asked to design (See Table 4.1, 'Neuron Assessment' column).

4.4.2 Comparison of the 'Neuron Assessment' Objectives with Those of Other Assessments

Experimental design difficulties were diagnosed using the RED (Dasgupta et al., 2014). The RED was developed and validated as a measure of five major areas of experimental design difficulties faced by undergraduate biology students. The areas are: the variable properties of an experimental subject; the manipulated variables; measurement of outcomes; account for variability; and the scope of inference appropriate

for experimental findings. The 'Neuron Assessment' as a diagnostic assessment, was designed to diagnose difficulties with the same RED areas. The 'Shrimp Assessment' (College Board 2006) and 'Drug Assessment' (SRI International, 2003) were published as valid measures of experimentation abilities for secondary school students, and in our previous report we found these measures to be useful indicators for knowledge and difficulties in RED areas for undergraduate biology students (Dasgupta et al., 2014). Specific RED areas and related concepts probed by each assessment are presented in Table 4.1.

Table 4.1: RED Areas Probed by Three Assessments

RED Areas and Concepts		'Shrimp Assessment'	'Drug Assessment'	'Neuron Assessment'
I.	**Variable Property of Experimental Subject**	Yes (a)*	Yes (a)	Yes (b)(c)
II.	**Manipulation of Variables**			
	a. Categorical (Discrete) Variable			Yes (b)
	b. Quantitative (Continuous) Variable			
	c. Treatment (Independent) Variable	Yes (a) (b)	Yes (a)	Yes (b)(c)(f)
	d. Control (Comparison) Group	Yes (a)	Yes (b)	Yes (c) (f)
	e. Combinatorial reasoning	Yes (a) (b)	No	Yes (c)
III.	**Measurement of Outcome**	No	Yes (e)	Yes (f)
IV.	**How to deal with variability with:**			
	a. Randomization of treatments	Yes (b)	Yes (d)	Yes (d)(h)
	b. Random (representative) sample	Yes (b)	Yes (c)	Yes (d) (h)
	c. Replication	Yes (a) (b)	No	Yes (h)
	d. Reducing effect of other variables	Yes (c)	Yes (c)	Yes (c)
V.	**Scope of Inference/Cause and Effect Conclusions/Interpretation of Findings**	Yes (c)(d)	Yes (f)	Yes (b)(e)(g)

* Letters in parentheses denote the assessment item sub question that probes for the given RED area concept. See Appendix L for details.

The 'Shrimp Assessment' requires students to design an experiment using different salt and nutrient levels to examine their effect on growth of tiger shrimps. Students are expected to present knowledge in following RED areas: They identify a variable property of the experimental subjects, manipulate appropriate variables in treatment groups, account for variability in their experimental procedures and estimate inferences from findings in their experiment (See Table 4.1, 'Shrimp Assessment' column). The 'Drug Assessment' tests the design of an experiment to develop a high blood pressure drug. According to the author (SRI International, 2003), this assessment prompts students to identify the variable property of the experimental subjects, organize control and treatments groups, suggest measures for experimental outcomes, propose ways to control variability and interpret findings from their experimental procedures (Table 4.1, 'Drug Assessment' column). The 'Shrimp Assessment' uses an everyday context with little scientific relevance while the 'Drug Assessment' has a social context of a blood pressure disorder. These two tests do not involve any explanations of biological mechanisms. The 'Neuron Assessment' presents a research problem in the

context of a neurological disease of impaired mitochondrial movement in neurons. Unlike the 'Drug' and 'Shrimp' assessment, this test challenges students to design an experiment that would yield information about a molecular mechanism.

4.4.3 Coding of Written Responses

Written responses to each of three assessments from four participants and the expert were inductively coded for evidence of knowledge of and difficulties with each of five RED areas. First, responses were examined for 'difficulties' according to 'Typical evidence of difficulties' in the RED (See Chapter 3, Appendix H). Next, if no difficulty was found, then 'correct' ideas were identified according to the propositional statements for each RED area. Finally, in case no difficulty or correct ideas were found, responses were coded as 'lack of evidence (LOE)' for a certain RED area. The goal of our coding task was to determine difficulties with each RED area and to see if knowledge or difficulty with a particular RED area shows up consistently across three different assessments.

4.5 Results

4.5.1 Students' Performance on the 'Neuron Assessment' in Comparison to Other Published Assessments

A comparison of students' written responses across three assessments showed that the 'Neuron Assessment' was capable of revealing knowledge and difficulties for five RED areas, similar to the 'Shrimp' and 'Drug Assessment'. Analyses of expert and student responses to the three assessments showed correct ideas for each RED component across three assessments. The correct ideas and difficulties revealed by the three assessments for the expert and each of four student participants summarized in Table 4.2 are discussed in more detail in the following section.

Table 4.2: Correct ideas and Difficulties with RED Areas Probed By three Assessments in written format.
Correct ideas and difficulties diagnosed by the 'Shrimp', 'Drug' and 'Neuron Assessment' for the five RED areas compared for expert and four student participants.

Areas of Experimental Design Difficulty	EXPERT		
	'Shrimp Assessment'	'Drug Assessment'	'Neuron Assessment'
Variable property of an experimental subject	Correct	Correct	Correct
Manipulation of variables	Correct	Correct	Correct
Measurement of outcome	Correct	Correct	Correct
Accounting for variability	Correct	Correct	Correct
Scope of inference	Correct	Correct	Correct

Areas of Experimental Design Difficulty	JUAN		
	'Shrimp Assessment'	'Drug Assessment'	'Neuron Assessment'
Variable property of an experimental subject	Difficulty (Subject considered variable)	Difficulty (Variable property diverges from study goal)	Difficulty (variable property diverges from study goal)
Manipulation of variables	Difficulty (Inappropriate treatment and control)	Difficulty (Controlling irrelevant variables)	Difficulty (Inappropriate control)
Measurement of outcome	Correct	Difficulty (No measures for outcome variables)	Difficulty (Outcome mismatches with investigation claim)
Accounting for variability	Difficulty (No randomization; No replication)	Difficulty (No randomization)	Difficulty (representative sample not considered; no replication)
Scope of inference	Difficulty (Overestimated inference)	Difficulty (Incorrect cause and effect relationship)	Difficulty (Incorrect cause and effect relationship)

Table 4.2 continued

Areas of Experimental Design Difficulty	EVE		
	'Shrimp Assessment'	'Drug Assessment'	'Neuron Assessment'
Variable property of an experimental subject	Correct	Correct	Correct
Manipulation of variables	Difficulty (No combinatorial reasoning)	Difficulty (Controlling irrelevant variables)	Difficulty (treatment vs. control group subjects not uniform)
Measurement of outcome	Correct	Difficulty (Mismatches with Instrument goal)	Difficulty (Mismatches with Instrument goal)
Accounting for variability	Difficulty (No randomization)	Correct	Difficulty (No randomization; treatment vs. control group subjects not uniform)
Scope of inference	Correct	Difficulty (Overstated inference)	Difficulty (Overstated inference)
Areas of Experimental Design Difficulty	LI NA		
	'Shrimp Assessment'	'Drug Assessment'	'Neuron Assessment'
Variable property of an experimental subject	Correct	Correct	Correct
Manipulation of variables	Correct	Correct	Correct
Measurement of outcome	Correct	Correct	Correct
Accounting for variability	Difficulty (No randomization)	Difficulty (No randomization)	Difficulty (No randomization)
Scope of inference	Difficulty (Overstated inference)	Difficulty (Overstated inference)	Difficulty (Overstated inference)

Table 4.2 continued

Areas of Experimental Design Difficulty	DANIEL		
	'Shrimp Assessment'	'Drug Assessment'	'Neuron Assessment'
(1a) Variable property of an experimental subject	Correct	Correct	Correct
(1b) Manipulation of variables	Difficulty (Haphazard treatment)	Difficulty (Haphazard treatment; No combinatorial reasoning)	Difficulty (Haphazard treatment; No combinatorial reasoning)
(1c) Measurement of outcome	Correct	Correct	Correct
(1d) Accounting for variability	Difficulty (Incomplete randomization; No replication)	Difficulty (Incomplete randomization; No replication)	Difficulty (Incomplete randomization; No replication)
(1e) Scope of inference	Correct	Correct	Correct

■ Correct Ideas ■ Difficulties ■ Lack of evidence
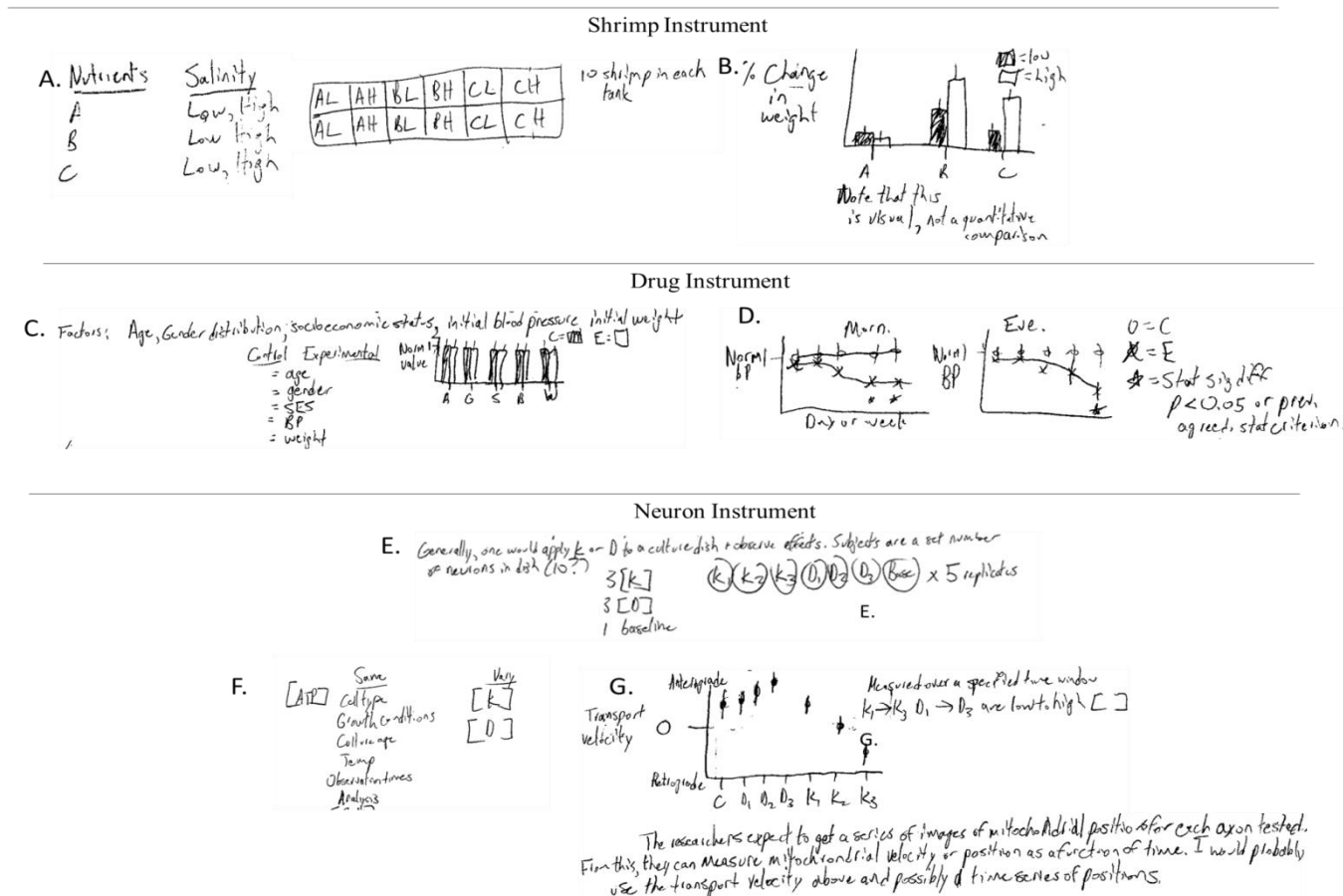
Figure 4.1: Expert's written assessment figure

Expert shows 'Shrimp Assessment' figures **A.** Tiger shrimp are considered to be the experimental subject and placed in treatment tanks with variable nutrient and salinity (*variable property of experimental subject*), Treatments are suitably assigned to individual tanks *(manipulation of variables)* **B.** The graph presents a causal relationship between % change in weight as an experimental outcome *(measurement of outcome)* based on the combined application of salinity and nutrients *(scope of inference)* and also shows errors bars to indicate variability *(accounting for variability)*; 'Drug Assessment' figures **C.** factors that are potential lurking variables are controlled between control and experimental group *(manipulation of variables. accounting for variability)*, **D.** The graph measures effect on blood pressure levels *(measurement of outcome)* in the control and experimental groups as a result of drug intake *(scope of inference)*; 'Neuron Assessment' figures **E.** Neurons are subjected to variable concentrations of compound K or D *(variable property of experimental subject, manipulation of variables)* along with replicates for each treatment *(accounting for variability)*; **F.** Non-relevant variables are maintained constant between various treatment groups *(accounting for variability)*, **G**. The graph presents a causal relationship between transport velocity of mitochondria *(measurement of outcome)* in axons in anterograde and retrograde directions under the influence of various concentrations of compound K and D *(scope of inference)*.

4.5.2 Experimental Design Knowledge and Difficulties Using RED across Three Assessments.

Expert.

Variable Property of Experimental Subject. In the 'Shrimp Assessment' the expert considered 'tiger shrimp' as the experimental subject, placed in treatment tanks with variable nutrient and salinity (Appendix L, Page 73 and Figure 4.1A). For the 'Drug Assessment', the variable property of blood pressure was reported *("Alamain will lower blood pressure in humans")*. For the 'Neuron Assessment', neurons were subjected to variable treatments of compound K or D (*"apply [K] or [D] to one culture dish with a set number of neurons";* Figure 4.1E).

Manipulation of Variables. The expert's 'Shrimp Assessment' response showed appropriately depicted treatment combinations of nutrients and sanity (Figure 4.1A). In the 'Drug Assessment', the expert correctly matched potential confounding variables across experimental groups (Figure 4.1C). For the 'Neuron Assessment', the expert maintained appropriate non-relevant variables constant across treatment and control groups (Figure 4.1F).

Measurement of Outcome. In the 'Shrimp Assessment', '% change in weight' of the shrimp was identified as the outcome variable (Figure 4.1B) while 'effect on blood pressure' (Figure 4.1D) was identified as the outcome for a graph plotted in the 'Drug Assessment'. The 'Neuron Assessment' considered 'transport velocity' of mitochondria as the outcome (Figure 4.1G).

Accounting for Variability. For the 'Shrimp Assessment', the expert correctly identified sources of variability *("tank temperature, measurement error, catch date")* and also used error bars to depict variability in outcome measures (Figure 4.1G). In the 'Drug Assessment', the expert recognized and controlled for variability introduced from potential lurking variables (Figure 4.1C). In the 'Neuron Assessment', the expert controlled for non-relevant factors as shown in Figure 4.1F.

Scope of Inference. In the 'Shrimp Assessment', the expert plotted a graph to present a causal relationship between % change in weight based on the combined application of salinity and nutrients (Figure 2B). The expert graphically represented inferences from the 'Drug Assessment' as changes in blood pressure levels vs. per day or week (Figure 4.1D). His 'Neuron Assessment' graph presented causal conclusions in the form of a graph showing relationship between transport velocity of mitochondria in axons in anterograde and retrograde directions under the influence of various concentrations of compound K and D. In summary, the expert demonstrated appropriate ideas across three assessments corresponding to all five areas of the RED (Table 4.2, 1a-e).

Figures drawn by the Expert (Figure 4.1 B, D and G) show both treatment and control group are represented side by side on the x-axis, appropriate outcome variables on the y-axis; errors bars represent variability of results from replication of treatments and a causal relationship can be coherently interpreted from a graphical representation. The expert figures were used to diagnose difficulties with RED areas presented by in figures created by students.

Shrimp Instrument

A. Tiger shrimps operate in the same way, they make use of the same nutrients and react the same way to all products. Having only tiger shrimps, makes it a controled factor

B.

control	only A	only B	only C	no nutrients only low salinity	no nutrients only high salinity	A + low salinity	A + high salinity	B + low salinity	B + high salinity	C + low salinity	C + high salinity

C. [graph: Length in cm vs variable factor A, B, C; legend: initial size of shrimp]

Drug Instrument

D. [bar graph: level of usage; legend: control, experimental; x-axis: diet exercise day, amount blood pressure, same level attention]

E. [bar graph: level of blood pressure in kPa; legend: control, experimental; x-axis: control experimental]

Neuron Instrument

F. [diagram: group with only K, group with only D, control — both compound K and D]

G. They will randomly ask equal amount of ppl with the same disorder and place them into each group.

The lack/substitution of compound K with a genetically engineered compound may lead to the patients getting better or vice versa. Same with compound D.

Figure 4.2: Juan's written assessment figures

Juan shows 'Shrimp Assessment' figures **A.** Experimental subject (tiger shrimp) are considered to be a control (*variable property of experimental subject*), no natural variability considered (*accounting for var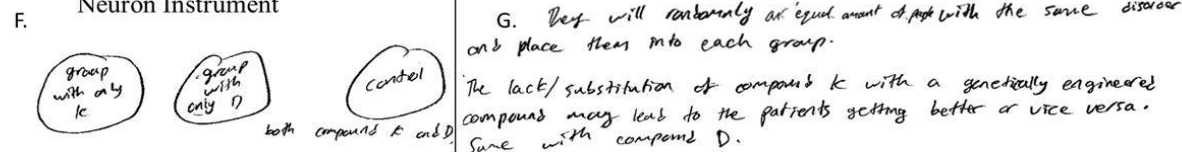iability*) **B.** Haphazard assignment of treatments, inappropriate controls with no nutrient or salinity (*manipulation of variables*), replication or randomization measures (*accounting for variability*); **C.** Graph with missing variability measures like error bars *(accounting for variability)*, flawed inferences as causal claims are inappropriate owing to haphazard treatments and missing variability measures *(scope of inference);* 'Drug Assessment' figures **D.** Graph y–axis reflects experimental subjects that diverge for the study goal (*variable property of experimental subject*), unrelated variables are matched in experimental groups (*manipulation of variables*), **E.** Graph y-axis shows no measures of stated outcome *(measurement of outcome),* no variability in the experiment (*accounting for variability*), owing to which flawed causal claims *(scope of inference)* are represented in this graph; 'Neuron Assessment' figures **F:** Experimental subjects (patients with disorder) diverge from the study goal (*variable property of experimental subject*), inappropriate controls (*manipulation of variables*), subjects who already carry disorder confounds causal claims to find disorder source and thus, represents an overestimated scope of inference *(scope of inference)*, **G:** No treatment replications are considered to improve experimental validity *(accounting for variability)*, outcome mismatches with investigation claim *(measurement of outcome)*.

Juan.

Variable Property of Experimental Subject. In the 'Shrimp Assessment', Juan erroneously considered the *experimental subject* (tiger shrimp) as part of the control (*"Having only tiger shrimps, makes it a controlled factor"*). In the 'Drug Assessment', he selected experimental subjects (Appendix L, Page 83) *"people with impaired kinesin, K and dynein, D")* confounding the experiment goal. In the 'Neuron Assessment' he also selected experimental subjects (*"people with impaired kinesin (k) and dynein (D)")* as confounders of the experiment goal.

Manipulation of Variables. For the 'Shrimp Assessment', Juan's treatments (*"tanks containing either a certain nutrient or salinity"*) were flawed because the treatments do not show any systematic combinatorial salinity and nutrients treatment combinations as required. For the 'Drug Assessment', he indicated unrelated variables (*"Patients are to be treated the same way; no individual attention"*). For the 'Neuron Assessment', his control groups (*"controls will carry compound K and D")* were flawed, as they do not carry normal cells required for disease diagnosis with provided materials as background for the assessment.

Measurement of Outcome. For the 'Shrimp Assessment', he suggested correct measureable outcomes, *"length in cm"* (Figure 4.2C). For the 'Drug Assessment', he proposed outcomes without specific measures on y-axis, *"level of blood pressure in 2 weeks"* (Figure 4.2E). For the 'Neuron Assessment', outcome variables ("*the lack/substitution of compound K with genetically engineered compound may lead to the patients getting better"*) mismatched the investigation claim to find a disorder source (Figure 4.2G).

Accounting for Variability. For the 'Shrimp Assessment', no natural variation among shrimp population was considered (*"Tiger shrimps operate the same way and react the same way to all products*") and variability measures like replication (using error bars) and randomization of treatments were missing in the graph (Fig. 4.2C). For the 'Drug Assessment', he controlled variability from variables that are unrelated to the study (*"Controlling working hours, exercise and administering the same amount of drug will*

eliminate the other variable factors"; Fig. 4.2D). For the 'Neuron Assessment', his variability measures were skewed because no treatment replications are considered to improve experimental validity (*"to improve the validity of the experiment the image software that measure mitochondrial movement in neurons will be used"*) (Appendix L, Page 83).

Scope of Inference. For the 'Shrimp Assessment', Juan overestimated experimental claims (Appendix L, Page 82), because his representations depict flawed treatments and missing variability measures (Fig. 4.2A, 4.2B) restrict inferences only to tiger shrimp. For the 'Drug Assessment', he suggested no causal relationships because flaws with outcome variables, error bars, representative sample, and randomization only suggest a correlational association. For the 'Neuron Assessment', Juan overestimated his findings as his experimental group subjects already carry the disorder (*"Researchers will randomly assign an equal number of people with the same disorder and place them into each group"*) (Fig. 4.2G). Experimental subjects who already carry disorder confound the derivation of causal claims.

Juan's graphical representations show difficulties with RED areas. His graph has no measures of stated outcome (Figure 4.2C), missing variability measures like error bars (Figure 4.2 C) and reflects experimental subjects that diverge for the study goal (Figure 4.2D).

In summary, Juan presented uniform difficulties reasoning with each of three published assessments in all RED areas. However, for *measurement of outcome,* he presented suitable outcome variables in the graph for 'Shrimp Assessment' but showed difficulty with graph for 'Drug Assessment' and explanations for 'Neuron Assessment' (Table 4.2, 2a-e).

Shrimp Instrument

A. shrimp — nutrient A, nutrient B, Nutrient C (x2 each)

B. In order to compare the growth, the average size of the shrimp must be known.

C. Tiger Shrimp — Other shrimp — Different Sizes, Different environments, Different results?

Drug Instrument

D. efficiency, the blood pressure of the participates should be recorded frequently. In order to judge the drug's

E. gender (♂ or ♀) + race (or or) + age range (20-30 or 31-40) + activity level

F. the new average is significantly lower than the original average, then the drug could be declared "successful".

G. In order to make the selection more "natural", the selections should be done at random. From there determine which will be in the control or experimental group. Separate draws can be done for both groups.

Neuron Instrument

H. healthy cell — Control — Experiment — cell with disorder — mitochondria

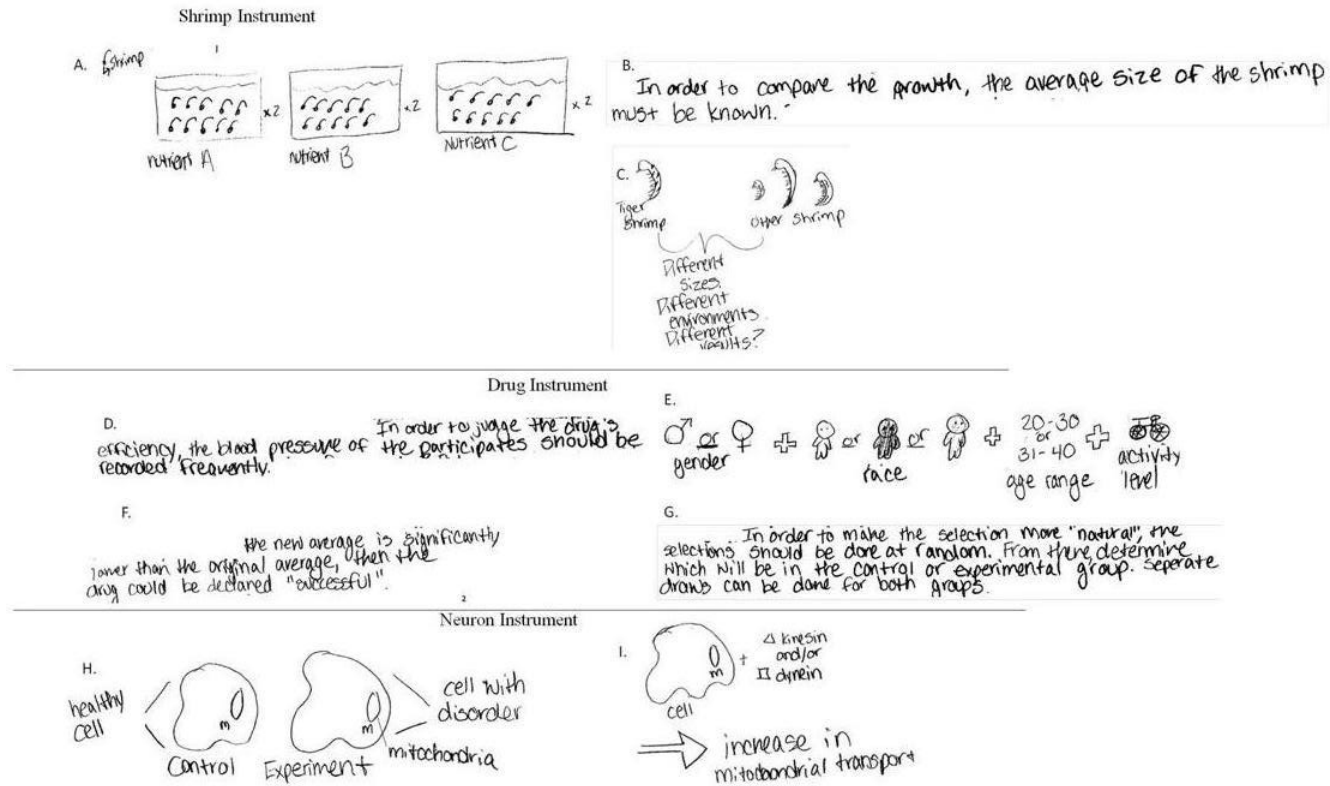I. cell + ▵ kinesin and/or Ⅱ dynein → increase in mitochondrial transport

Figure 4.3: Eve's written assessment figures

Eve shows 'Shrimp Assessment' figures **A.** Tiger shrimp are suitably placed in variable treatment tanks *(variable property of experimental subject);* Missing combinations of nutrients and salts as treatment variables (*manipulation of variables*) and missing randomization measures (*accounting for variability*); **B.** Appropriate outcome measures; **C.** Suitable estimation that using only 'tiger shrimp' limits scope of inference of experimental findings *(scope of inference).*'Drug Assessment' figures **D.** Appropriate variable property "blood pressure" is considered for experimental subjects *(variable property of experimental subject)* ; **E.** unrelated variables like 'limited age range' are matched across experimental groups *(manipulation of variables*), inappropriate inferences are made owing to biased sorting of subjects *(scope of inference)* ; **F.** proposed outcomes mismatches the hypothesis of reduced blood pressure*(measurement of outcome)*; **G.** variability measures by randomly assigning participants to experimental groups *(accounting for variability)*'Neuron Assessment' figures **G.** "Cell" is a suitable experimental subject with "disorder" as variable property *(variable property of experimental subject)* but variables in treatment vs. control group subjects are not uniform *(manipulation of variables)* and biased subjects do not indicate a representative sample for the study *(accounting for variability)*; **H.** The outcome 'increase in mitochondrial transport' mismatches with the assessment goal *(measurement of outcome)* and depicts a causal pattern that mismatches with the given background information as kinesin and dynein which are mitochondrial transport inhibitors are shown to increase transport *(scope of inference)*.

Eve.

Variable Property of Experimental Subject. For the 'Shrimp Assessment', Eve identified an appropriate experimental subject with a variable property *("Biologists intend to use tiger shrimp to compare their growth to test 3 different growth enhancing nutrients and 2 salinity levels"* (Appendix L, Page 89). For the 'Drug Assessment', she considered experimental subjects with an appropriate variable property (*"blood pressure"*) (Figure 4.3D). For the 'Neuron Assessment', Eve correctly identified an experimental subject as evident from her diagram (Fig. 4.3H, see mitochondria).

Manipulation of Variables. For the 'Shrimp Assessment', by ignoring salinity, Eve failed to combine two treatments as required to address the experimental goal (Fig. 4.3A). For the 'Drug Assessment', she controlled for unrelated variables like gender, race and age in the study (Fig. 4.3E). For the 'Neuron Assessment', her visual representation (Fig. 4.3H) showed biased selection of control vs. treatments group participants.

Measurement of Outcome. For the 'Shrimp Assessment', she presented measurable outcomes *("with the average size of shrimp recorded, the results of the other tanks have a basic unit for comparison")*. For the 'Drug Assessment', outcome variables *("Significantly lower blood pressure with the drug")* did not match the given experimental goal as only lowering blood would deem the drug effective whether statistically significant or not. For the 'Neuron Assessment', the identified outcome variables *("healthy amount of mitochondrial movement deems the experiment successful")* did not match the experimental goal of finding the source of the mitochondrial disorder.

Accounting for Variability. In the 'Shrimp Assessment', Eve failed to deal with variability by considering randomization of treatments to the shrimp tanks (*"Remove shrimp from each of 12 tanks and record their growth"*). For the 'Drug Assessment', she accounted for variability by randomly assigning participants to experimental groups (*"participant assignment for control or experimental group should be done at random"*). For the 'Neuron Assessment', she assigned cells to control and experiment group with no

information about controlling variability, using measures like randomization. Her selection of control and experimental group subjects were also biased as she suggested using a healthy and disordered cell for the same treatments (Figure 4.3H).

Scope of Inference. For the 'Shrimp Assessment', experimental inferences that were overstated with flawed treatment combinations and no consideration for variability measures like randomization (Fig. 4.3A). For the 'Drug Assessment', experimental subject in treatment vs. control groups were biased as they were group according to an unrelated variable likes race (Figure 4.3E) which resulted in flawed causal inferences. For the 'Neuron Assessment', causal claims were flawed since she explained a causation pattern *("kinesin inhibitor increase mitochondrial movement")* mismatched with assessment background which states kinesin allows mitochondrial movement and its inhibition will decrease and not increase movement.

Eve shows abilities to construct a representation to illustrate organization of experimental variables for all three assessments. For the 'Shrimp Assessment", she was able to construct treatment tanks with shrimp (Figure 4.3 A). For the 'Drug Assessment', she depicted biased manipulation of variables unrelated to the context (Figure 4.3E) and for the 'Neuron Assessment', she was able to illustrate the treatment, control (Figure 4.3H) and outcome variables as considered in her explanations (Figure 4.3I).

In summary, Eve presented correct ideas about the *variable property of experimental subject* but showed trouble with *manipulation of variables* and *scope of inference* across three assessments. She struggled with *measurement of outcome* with the 'Drug' and 'Neuron Assessment' as found with outcome variables identified in her graph. She showed flaws in *accounting for variability* with the 'Shrimp' and 'Neuron Assessment'. Her visuals for the 'Neuron Instrument' depicted biased sorting of treatment and control group subjects. Overall, Eve shows correct ideas for RED areas like *measurement of outcome* for 'Shrimp Assessment' and *accounting for variability* for 'Drug Assessment' but faced difficulty with these areas with other assessments (*measurement of outcome* was difficult for 'Drug' and 'Neuron' assessment and
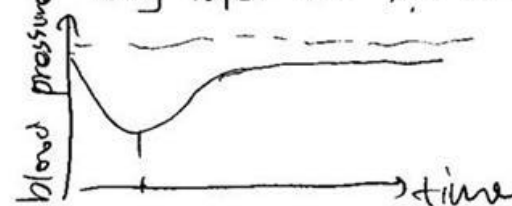
accounting for variability for difficult for 'Shrimp' and 'Neuron' assessment) (Table 4.2; 3a-e).
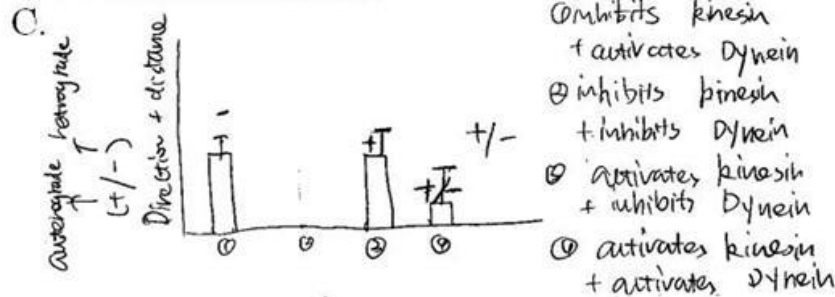
Shrimp Instrument

A.
- A + low → group 1 and 2
- A + high → group 3 and 4
- B + low → group 5 and 6
- B + high → group 7 and 8
- C + low → group 9 and 10
- C + high → group 11 and 12

Drug Instrument

B. measure the blood pressure of the participants within the efficient time (the same). they take the Alamain

blood pressure ↑ → time

Neuron Instrument

C. Anterograde ← (+/-) → retrograde / Direction + distance

① inhibits kinesin + activates Dynein
② inhibits kinesin + inhibits Dynein
③ activates kinesin + inhibits Dynein
④ activates kinesin + activates Dynein

+ − +/− +/−
① ② ③ ④

D.

Dynein ← → kinesin
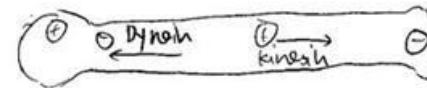
Figure 4.4: Li Na's written assessment figures

Li Na shows 'Shrimp Assessment' figures **A.** 'Tiger shrimp' are appropriately placed in variable treatment tanks *(variable property of experimental subject)*, treatment variables are manipulated and organized in given treatment tanks *(manipulation of variables)*; 'Drug Assessment' figures **B.** suitable knowledge of causal explanations along with visualizations to indicate "lower blood pressure with use of drug Alamain" 'Neuron Assessment' figures **C.** suitable outcome measures "direction and distance of movement" are considered on y-axis *(measurement of outcome)* and represented causal pattern (*"kinesin causes anterograde movement and dynein causes retrograde movement due to ion interactions"*) do not match with given experimental goal *(scope of inference)*.

Li Na.

Variable Property of Experimental Subject. In the 'Shrimp Assessment', Li Na identified appropriate experimental subjects with a variable property *("Treat the shrimp with salinity and growth enhancing nutrients to see effects on growth"*, Appendix L Page 97*).* For the 'Drug Assessment', she also showed correct ideas about experimental subjects *("Alamain can lower the human blood pressure").* For the 'Neuron Assessment', however, she suggested a variable property inconsistent with the experiment goal (*"different concentration of motor proteins and ATP might affect movement of mitochondria").*

Manipulation of Variables. In the 'Shrimp Assessment', she represented the correct combination of treatments suitable to the experiment goal (Fig. 4.4A). For the 'Drug Assessment', however, she considered irrelevant variables *("participants with same age, gender, nationality"*; Appendix L, Page 99). For the 'Neuron Assessment' she also controlled for irrelevant outside variables, *"Concentrations of cellular complexes, same diffusion pressure"* (Appendix L, Page 102*).*

Measurement of Outcome. In the 'Shrimp Assessment', she suggested correct measurable outcomes *("compare body length of shrimp in three weeks".* For the 'Drug Assessment', she depicted correct measureable outcomes (*"lowered blood pressure"*) in her graph (Fig. 4.4B). For the 'Neuron Assessment', she identified correct measureable outcomes such a "direction of movement" on the graph y-axis (Fig. 4.4C).

Accounting for Variability. In the 'Shrimp Assessment', she showed difficulties dealing with variability, as her experimental subjects were not representative of the targeted shrimp population ("*shrimps should be similar in gender"*, Appendix L Page 99). For the 'Drug Assessment', she dealt with variability suitably, considering measures like randomization (*"Participants in experimental groups may be assigned in a randomized block experiment").* For the 'Neuron Assessment', her experiment did not show any variability measures as in response to a probe about assignment of subjects, she draw a graph representing only treatment but provided no explanation about randomizing treatments (Figure 4.4C).

Scope of Inference. For the 'Shrimp Assessment', Li Na appropriately estimated the scope of inferences *("Only tiger shrimps can't be representative of all shrimps")*. For the 'Drug Assessment', she presented correct causal explanations (*"Measure blood pressure in a given time period to see the efficiency of Alamain")* and supporting graphical representations (Fig. 4.4B). For the 'Neuron Assessment' her graphical results (Fig. 4.4C) represented a causal pattern (*"kinesin causes anterograde movement and dynein causes retrograde movement due to ion interactions")* not matched to the given experimental goal, which was to understand effect of kinesin or dynein inhibition on mitochondrial movement .

Li Na's figures for the three assessments show knowledge and difficulties with RED areas. She represented manipulated treatment variables organized in given treatment tanks (Figure 4.4 A), her graph depicts a causal relationship (Figure 4.4 B) and suitable outcome measures are considered on the graph y-axis (Figure 4.4C). In summary, certain RED areas like *Accounting for Variability* and *Scope of Inference* were diagnosed as difficulties for all three assessment contexts in the case of Li Na. For some RED areas including, *manipulation of variables* where she presents correct ideas with the 'Shrimp Assessment' but struggles with the 'Drug' and 'Neuron Assessment' contexts (Table 4.2; 4a-e).
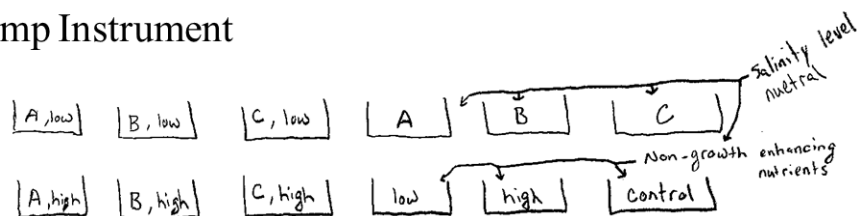
Shrimp Instrument

Figure 4.5: Daniel's written assessment figures
Daniel's figure shows 'Shrimp Assessment' figures where **A.** Tiger shrimp appropriately placed in variable treatments *(variable property of experimental subject)*, treatment assignment to tanks are haphazard *(manipulation of variables)* and no replication measures *(accounting for variability)*

Daniel.

Variable Property of Experimental Subject. Under the 'Shrimp Assessment', Daniel portrayed correct knowledge in this area ("*tanks will combine a nutrient and salinity level to determine how each effect shrimp's growth*") (Figure 4.5A). Under the 'Drug Assessment', Daniel considered the experimental subject correctly *("Alamain given to patients…their blood pressure will lower")*. For the 'Neuron Assessment', we found no difficulties with this area ("*vary treatment to determine mitochondrial movement changes*").

Manipulation of Variables. In the 'Shrimp Assessment', Daniel indicated difficulty with variables due to inappropriate treatment combinations (Figure 4.5A). For the 'Drug Assessment', he considered irrelevant variables *("same doctor across experimental groups")* for his study. For the 'Neuron Assessment', he described this area accurately (*"same axon and environment controls for mitochondria to be maintained; treatments to be varied"*).

Measurement of Outcome. For the 'Shrimp Assessment', he suggested suitable outcome measures (*"To determine growth length and mass of the shrimp should be measured")*. For the 'Drug Assessment', he considered irrelevant measurable outcomes ("*Diet and stress level")*. For the 'Neuron Assessment', he suggested appropriate outcomes ("*image software determines change in mitochondrial movement")*

Accounting for Variability. In the 'Shrimp Assessment', his considerations for dealing with variability shows difficulties as treatment assignments showed no replications (Figure 4.5A) and incomplete randomization (*"Randomly assign shrimp to each tank"*). In the 'Drug Assessment', he made appropriate variability considerations like randomization *("put patients in groups using a random number generator")*. In the 'Neuron Assessment', he accounted for good variability measures *("randomly picked and assign participants to experimental groups")*.

Scope of Inference. In the 'Shrimp Assessment', he appropriately explained experimental inferences ("*Different shrimp may grow better or worse under different conditions, meaning for the results to apply to all shrimp a study on each species must be done"*; Appendix L, Page 106). In the 'Drug Assessment', he drew plausible causal interpretation for this study in this assessment *("Experimental group will have an average larger drop in blood pressure than the control group"*; Appendix L, Page 108). In the 'Neuron Assessment', he proposed appropriate experimental inferences ("*treatment with kinesin may stop movement towards terminal and away from terminal for dynein"*; Page 110).

Daniel created a representation showing treatment groups for the 'Shrimp Assessment' which revealed difficulties as treatments were haphazardly applied to the tanks and were not replicated (Figure 4.5A). Interestingly, while the prompts specifically asked students to create visualizations to depict experimental knowledge, Daniel only drew a figure for the ''Shrimp Assessment' and provided written explanations for the RED areas with other assessments. Thus Daniel shows no difficulties with the *variable property of experimental subject* in all three assessments. But struggled with *manipulating variables* in the 'Shrimp' and 'Drug Assessment', with *measurement of outcome* in the 'Drug Assessment' and *accounting for variability* in the 'Shrimp Assessment' (Table 4.2; 5a-e).

4.6 <u>Discussion</u>

This study addresses the research question how well does students' performance on the 'Neuron Assessment' compare with their knowledge and difficulties revealed by other assessments of experimental design knowledge in biology. The 'Neuron Assessment' was found to be a comparable measure of experimental design, with published 'Shrimp' and 'Drug' assessment as indicated by expert and student responses across the three assessments. The other assessments have been published previously and for use with secondary school students and these have been shown as useful measures of experimental design abilities with our own undergraduate students as well (Dasgupta et al., 2014).

Taken together, our data suggest that the 'Neuron Assessment' is good diagnostic assessment because the expert was able to present accurate ideas with 'Neuron Assessment' that were similar in nature to his responses and diagrams depicted for the two published assessments. Students who showed knowledge on the published assessments also showed knowledge with the 'Neuron Assessment' with very few exceptions. Similarly, students who performed poorly on the 'Neuron Assessment' also performed poorly on the other assessments. This means that the expert was able to apply experimental design knowledge in a domain general manner across the three assessment contexts (Feltovich, Prietula, and Ericsson, 2006; Chi, Feltovich, and Glaser, 1981; Ericsson, 2006). Students' knowledge and difficulties revealed with the 'Neuron Assessment' was comparable to the published assessments. This could be because the 'Neuron Assessment' background provided domain specific information to help students present their domain general knowledge across three assessment contexts.

There is strong evidence that 'Neuron Assessment' is equally valid as a measure of the knowledge areas characterized in the RED as the two published assessments that were used to develop the RED. Knowledge and difficulties detected coincided with all five areas across four students with at least one published assessment. For 16 of 20 areas knowledge or difficulties detected were identical across all three assessments. In 4 areas results with the 'Neuron Assessment' were identical to one other assessment but the third

published assessment differed. In fact where students had difficulty with the 'Neuron Assessment', that difficulty was reflected in their difficulty for creating a visual representation of the experiment.

Students got an opportunity to present their experimental ideas in written as well as oral format for the 'Neuron Assessment' (Chapter 3 and 4). Thus, it is interesting that Daniel who showed difficulties with manipulation of variables and accounting for variability across all three assessments in written format, actually showed no difficulties as he orally explained his ideas for an experiment with the 'Neuron Assessment'. In fact if comparing Daniel's performance in the oral interview with his written responses indicates how his written responses may not present a complete representation of his knowledge in general. However, this poses as a problem with written format assessments information is limited to only what student choose to write. The oral interviews were carried out once the written assessment responses were turned in.

Only in a few cases do we find failure with transfer of knowledge, meaning that certain RED areas reveal variable knowledge across three assessments. For example, Juan showed correct knowledge for *measurement of outcome* in the 'Shrimp Assessment', but showed difficulties in the 'Drug' and 'Neuron' assessments while Eve showed correct knowledge for *accounting for variability* for the 'Drug Assessment' but flawed ones with the other two assessments. Thus, variable knowledge for a certain RED area across three assessments helped us realize that this may indicate an area of knowledge development for that student. In other cases where a student struggles with a RED area across all three assessments, we can be certain that they have not yet developed any knowledge of the area. However, if they carry knowledge but the assessment context leads them astray, then they may show correct ideas for certain assessments but not others (Detterman & Sternberg, 1993). This should not be the case if they have developed domain general reasoning abilities as was demonstrated by the expert (Zimmerman 2000, 2007).

More difficulties with RED areas were detected with the 'Neuron Assessment' responses than with other assessments. This shows that either students lack correct knowledge of experimental concepts and visualizations in general, or the context of the

'Neuron Assessment' makes it difficult for them to apply correct knowledge to reason about experimental design. It has been shown that solving problems in a rich knowledge domain is often easier for experts than for novice students because experts tend to categorize problems better and use specific principles based on their own knowledge of how to solve the problem (Chi, Feltovich, and Glaser, 1981). It could also be that only the 'Neuron Assessment' required "mechanistic reasoning" (refers to description of a biological mechanism about how the component entities of a biological phenomenon interact at the molecular, microscopic, and macroscopic levels to produce detectable changes in state, activities, and spatial and temporal organization) which may be a domain-general skill not yet developed by students who performed better on the 'Shrimp' or 'Drug' assessment than the 'Neuron Assessment'. However, this study shows that none of the students performed better on one assessment than the other.

Our findings demonstrate that the expert showed correct knowledge of five RED areas in all assessment contexts. As a neurobiologist, the expert showed knowledge of RED areas in the 'Neuron Assessment' just as well as for the 'Shrimp' and 'Drug' Assessments. Previous literature reports that experts derive cues from the domain of a given problem based on their own knowledge of the field, but an important question remains whether experts can similarly gather cues in knowledge domains that don't belong to their expertise areas (Chi, Feltovich, and Glaser, 1981).

This study presents an original, 'Neuron Assessment' based on current research that is shown to be as comparable to other published assessments, a valid and useful measure of five areas of experimental design based on the RED. As a unique aspect, the assessment levels for all prior knowledge differences by providing all required background and visualizations required to design an experiment involving a mitochondrial movement disorder in neurons. Examination of knowledge and difficulties across RED areas illustrates very little evidence of problems with transfer because in contrast to the expert, students struggle to apply knowledge presented in one assessment domain to other assessment designed in completely different domains. This indicates that knowledge of RED areas are perhaps dependent on the context and complete

understanding of student ideas about experiments requires testing of their abilities across multiple contexts.

References

Barnett SM, Ceci SJ (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological bulletin *128*, 612.

Chen Z, Klahr D (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. Child Dev *70*, 1098-1120.

Chi MT, Feltovich PJ, Glaser R (1981). Categorization and representation of physics problems by experts and novices. Cog Sci *5*, 121-152.

Dasgupta AP, Anderson TR, Pelaez N (2014). Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties. CBE-Life Sci Educ *13*, 265-284.

Detterman DK, Sternberg RJ (1993). Transfer on trial: Intelligence, cognition and instruction, p. 99.

Ericsson (2006). The influence of experience and deliberate practice on the development of superior expert performance. The Cambridge handbook of expertise and expert performance, 683-703.

Feltovich PJ, Prietula M J, Ericsson, KA (2006). Studies of Expertise from Psychological Perspectives.

Zimmerman C (2000). The development of scientific reasoning skills. Developmental Review *20*, p. 102.

Zimmerman C (2007). The development of scientific thinking skills in elementary and middle school. Developmental Review *27*, 172-223.

CHAPTER 5: CONCLUSIONS

5.1. <u>Dissertation Focus</u>

This dissertation provides and evaluates tools to address foster scientific thinking, in particular, experimental design competencies which are critical to undergraduate biology education. Several research calls highlight the necessity to increase understanding of the experimental research process as a core scientific ability (for e.g., AAAS, 2011; AAMC-HHMI, 2009; NRC, 2007). To do so, effective assessments are required to ascertain the scientific knowledge that students actually possess and are able to demonstrate for designing experiments.

Students can acquire subject matter knowledge by evaluating experimental evidence in biology courses. But changes in knowledge can only be identified by actually measuring what students learn about experiments. A well-designed course carries tight alignments across learning objectives, instruction and assessment. Reliable assessment tools play an important role in course instruction because they provide a clear idea about students' difficulties which instructors can use to target remediation strategies. In this regard, assessments that help faculty and students diagnose experimental design abilities can allow identification of activities to best promote these abilities. This dissertation describes a range of qualitative and quantitative approaches for instructors to diagnose difficulties with design and visualization of experiments faced by undergraduate biology students.

5.2 <u>Summary of Dissertation Chapter Findings</u>

Chapter 2 of this dissertation presents development of a <u>R</u>ubric of <u>E</u>xperimental <u>D</u>esign (RED) which identifies five major areas of undergraduate biology students'

difficulties with experimental design in biology. *Chapter 3* describes development and testing of an original 'Neuron Assessment' based on current research. A case study method was conducted with oral interviews to investigate interactions among three factors, conceptual knowledge (C), reasoning skills (R) and modes of representation (M). This chapter characterizes expert ways of designing an experiment and examines how useful the assessment is to distinguish students who can do the same from those who have difficulty designing experiments. *Chapter 4* of this dissertation compares the 'Neuron Assessment' responses with two published assessments that target similar knowledge and difficulties with experimental design. A case study method is used to gather responses in a paper pencil format test. All three assessments are compared in terms of how well each probes for the RED areas. Findings show to what degree each assessment context reveals different difficulties. In the current chapter (*Chapter 5*) we highlight major findings from each study, the implications of this dissertation in the realm of experimental design and biology education research, and propose future avenues of research to further extend the findings from this work.

Chapter 2 (Design and development of the RED; Dasgupta et al., 2014) investigates student knowledge and difficulties with experimental design. *Established* difficulties were identified in three or more studies, found in two or more populations and carried enough prevalence in data to support a stable description of a difficulty according to our literature review. All *Established* difficulties were consistently found in responses from our own undergraduate biology students. Data from our undergraduate biology students permitted the re-classification of one *Partially Established* difficulty, the *variable property of experimental subject*, to *Established*. Data collected from undergraduate biology students, together with difficulty data from a review of the literature, confirmed five major areas of difficulty with experimental design: *(1) a property of an experimental subject that is variable; (2) manipulation of variables; (3) measurement of outcome; (4) accounting for and measuring variability and (5) scope of inference of findings*. Data from three assessments was used to inform the development of a Rubric for Experimental Design (RED), consisting of descriptions of correct ideas and typical difficulties within each of the abovementioned 5 major areas. The RED was

shown to be an effective tool for detecting changes in undergraduate students' experimental design knowledge during instruction. The study design for this chapter is more like a post-test only kind of design as some students took a course while and others did not yet have a course. Thus, there could be all kinds of interference and it cannot be particularly inferred that anything differences in knowledge observed were specifically as a result of taking the course. Thus, whether experimental design was taught explicitly or not is irrelevant but we now know what should be taught for students like those enrolled in that course.

The RED shows great promise for diagnosing students' experimental design knowledge in lecture settings, laboratory courses, research internships and Course-based Undergraduate Research Experiences (CUREs). It also shows potential for guiding the development and selection of assessment and instructional activities to do with experimental design.

Chapter 3 (Development and testing of the 'Neuron Assessment') described the development and usefulness of an originally designed 'Neuron Assessment' based on a cutting edge research context to examine student abilities with visualizations important to experimental design, using the concept-Reasoning-Mode of Representation (CRM) model (Schönborn and Anderson, 2009). Findings related to visual (RM) abilities showed that before the 'Neuron Assessment', the expert presented suitable visualizations of mitochondrial movement along neurons but showed no experimental design visuals like experimental tables or a graph with comparison groups. With the assessment, the expert interpreted the provided figures and created appropriate visual representations with experimental tables and graphs appropriate for investigation with mitochondrial movement. Examination of experts' conceptual reasoning (RC) abilities before the 'Neuron Assessment', revealed use of mitochondrial movement but no knowledge of treatment and control variables in the experiment designed to test mitochondrial disorder in neurons. The 'Neuron Assessment' was a good probe because it prompted the expert to propose an experiment that carried suitable knowledge of all areas of the RED.

The 'Neuron Assessment' was applied across a range of students and Juan and Eve were found to be the more typical performing students while Li Na and Daniel were more "expert-like" students. Findings from participants' visual (RM) abilities showed that the 'Neuron Assessment' background and provided figures were decoded by the expert and students alike. Everybody constructed representations before and with the 'Neuron Assessment'. All students were able to use the 'Neuron Assessment' and horizontally translate between experimental tables and graphs. The expert interpreted temporal resolution before and with the 'Neuron Assessment' but students represented comparison groups with the assessment information. However, the 'Neuron Assessment' did not probe sufficiently for visualizing levels of organization for the expert and Li Na and Daniel. Both experts and students showed no evidence of spatial manipulation abilities with the 'Neuron Assessment'.

Findings from conceptual reasoning (RC) abilities showed that all participants had knowledge of 'neurons' before and with the 'Neuron Assessment'. All students also knew about 'organelle movement' before and after the assessment except one student who showed this knowledge only after being given the assessment. 'Experimental subject' and 'variables' were considered by all participants before and with the 'Neuron Assessment' but Eve had trouble with presenting this knowledge for the 'Neuron Assessment'. Surprisingly, the expert provided no knowledge of 'treatment variables' before the assessment but explained this with the 'Neuron Assessment'. Juan and Eve faced difficulties proposing treatment variables with the assessment. With 'control variables' Li Na showed evidence of difficulty only with the 'Neuron Assessment'. Expert and three students (except Juan) were able to identify appropriate 'confounding variables' with 'Neuron Assessment'. All participants identified 'outcome variables' regardless of the assessment. In terms of 'variability' in an experiment, the expert and Daniel demonstrated this knowledge before and with 'Neuron Assessment' while Juan and Eve didn't consider variability at all and Li Na discussed it only with the assessment. Finally, cause and effect explanations were considered by the expert both before and after but Juan and Eve showed this knowledge only with 'Neuron Assessment'.

Chapter 4 presents an account of the expert and student written responses for 'Neuron Assessment' in comparison with two published 'Shrimp' and 'Drug' assessment. Comparable difficulties were found with RED areas (Dasgupta et al., 2014) on all three assessments with only a few exceptions. Interestingly, for a particular RED area, some students who presented correct knowledge with a certain assessment but struggled with others. For example, Eve had correct ideas for measurement of outcome for the 'Shrimp Assessment' but showed difficulties with the 'Drug' and 'Neuron' assessment. This alludes that students' reasoning about RED areas are perhaps dependent on context, as the three assessments presented variable backgrounds. It could also be that the 'Neuron Assessment' is the only one of three that requires a mechanistic explanation which is perhaps a domain general skill for some but we do not yet know.

5.3 Research Implications of This Dissertation

The findings established in the studies of this dissertation hold broader implications for both theory and practice. First, as an original contribution, the second chapter of this dissertation presents and validates a Rubric for Experimental Design (RED) that characterizes five major areas of experimental design difficulties faced by undergraduate students. A broad implication of the RED is its role as a tool to identify students' experimental design deficiencies. Information about specific difficulties might perhaps reveal a need to formulate new learning objectives along with activities and remediation strategies to fix such deficiencies and difficulties. The RED can be applied towards designing instructional strategies to alert both students and instructors as to pitfalls to avoid and areas in need of instruction to promote proficiency with experimental design.

Usefulness of RED: The RED has potential to be useful for measuring progress from experiential learning with laboratory courses, research internships, or Course-based Undergraduate Research Experiences (CUREs) (Auchincloss et al., 2014) and not just with lecture courses as exemplified in Dasgupta et al., 2014. According to Laursen et al. (2010), undergraduate research experiences are often evaluated by faculty, and some *"ask students to 'demonstrate their understanding of the processes of science' by framing*

a research question, developing a hypothesis, designing an experiment to test it, analyzing real data, writing a research report, and presenting their own work. These examples were sparse, and institutional evaluation efforts were often described as poorly developed or even perfunctory" (p. 176). The RED may serve as a useful guide for assessing assignments to help students develop experimental design abilities by faculty mentors who consider the various research contexts appropriate for their local situation. As a standard rubric, the RED may be useful to draw interpretations from other assessments of students' abilities to design experiments. The RED helped us find information about areas where our own students needed assistance as we strove to teach students not just knowledge of the subject matter but how biology is performed as a research endeavor. The application of RED could be useful at all stages of learning, including objectives, instruction and assessment of experimental design. In fact the RED informed diagnosis of knowledge and difficulties in response to an assessment presented in *Chapter 4*, and it could be useful for faculty who want to generate more assessment items as described here in *Chapter 3*.

Usefulness of 'Neuron Assessment': Second, development and testing of the original, 'Neuron Assessment' provides a probe that can be used to test student abilities to reason about visual representation of their experimental design knowledge. Comparison of 'Neuron Assessment' responses with other published assessments yields differences for the same experimental design abilities when tested across different contexts. We found that the 'Neuron Assessment' revealed difficulties with certain RED areas which are different than difficulties revealed by the 'Shrimp' and 'Drug' assessment and vice versa. As example, two students showed difficulty with measurement of outcome in the 'Drug' and 'Neuron' Assessment, but correct ideas in 'Shrimp Assessment' (Refer to Chapter 4, Table 4.2). This suggests that multiple assessments based on different contexts might be used in combination in order to get a better idea about student difficulties with a certain RED component. In other words, to confirm whether the difficulty lies with a certain RED area or related to context of the assessment, it will be useful to measure the same area using more than one assessment. It

is important to triangulate students' experimental design abilities and difficulties in a range of contexts, because it might be that a certain context/domain leads them astray.

The 'Neuron Assessment' bears future research implications as it guides development of new diagnostic assessments in other biology subject areas. The assessment redundantly provides the same contextual knowledge in multiple modes of written text and visualizations. This is useful as the only way to test for domain general knowledge is to provide the domain specific knowledge about a context belonging to the biology subject matter areas. This 'Neuron Assessment' was successful at targeting experimental design concepts (Chapter 3) and RED areas comparably to the 'Shrimp' and 'Drug' assessment (Chapter 4). Similar formats can be used to design assessments based on other biology topics like cell biology, ecology, or genetics. Background information presented with visuals is useful to level prior knowledge differences and thus analysis of responses can be easier as the focus can be examining abilities to design an original experimental investigation.

Knowledge gained from research reported here has already been applied in several ways. First, experimental design based teaching modules was designed to help biochemistry faculty with pre med undergraduate students. Second, experimental design learning objectives were developed and assessed in a large enrollment introductory biology course. Third, assessment design for a course based research project was carried out as an external evaluator.

Design of Modules to Test Experimental Design: Practice gained with creating experimental design assessments was applied towards supporting biochemistry faculty in the design new modules such as 'Detection of colon cancer via PCR of feces' for pre-med undergraduate level students. An existing module was modified to provide students with visualizations that depict double-stranded DNA and the amplified gene sequence so that the activity would focus students on research and not just content. A question regarding a reasoning about PCR experiment was reorganized in order to have students examine and select appropriate examples from a pool of previous student responses with correct ideas and difficulties and to provide justifications for their reasoning.

Assessing Students' Learning about Biological Experiments in a Large Enrollment Lecture: A study to investigate connections between student perceptions about experimental reasoning and biology subject matter was carried out using student ratings for a self-reported questionnaire, the perception inventory (PPI) in an undergraduate first year biology course (Clase et al., 2010). Summative assessment items were designed to measure the effect of course-based research on student learning and attitudes. Summative assessment used a Participant Perception Inventory or PPI (Clase et al., 2010). A PPI consists of survey items designed to quantify student responses in the dimensions of knowledge, experience, and confidence. Students were asked to indicate their perceptions of knowledge (K), experience (E), and confidence (C) on a low to high (1-5) Likert scale for each of 30 learning outcome statements in six categories. The PPI was developed to track target course outcomes in six potentially overlapping biology sub-disciplines: the *physical and chemical basis of life, molecular basis of regulation, plant biology, animal biology,* and the *experimental* and *empirical basis of biology* (Appendix M). For each item, students' reported KEC were averaged to yield one score per student. Descriptive statistical methods were applied to study variations in student ratings to reveal the different clusters of biology knowledge that represent groups of learning outcomes in the PPI that vary together. Pre- and post-instruction mean values were calculated for each item and averaged to obtain overall category means for each of the six PPI categories (Appendix N). Subsequently, we used factor analysis to examine connections and separations established from the variations in ratings between

experimental biology and other biology subject areas which were represented visually with the help of a network diagram using innovative PAJEK software (Appendix O).

Average pre and post instruction KEC scores for each learning outcomes across six categories were listed by increasing order (Appendix N). Pre-instruction category KEC means are lowest for *'Molecular Basis of Regulation'* and highest for *'Experimental Design'*. In contrast, the post-instruction category KEC means are relatively close, ranging from 3.52 to 3.76 with the exception of *'Experimental Design'* that yield a higher category mean of 4.14 (Appendix N). This indicates that students were not as aware of their deficient knowledge of experimental design compared with *'Molecular Basis of Regulation'*. Factor Analysis validated the target subject areas identified for the course. A correlation network diagram energized using the Kamada-Kawai transformation (Appendix O), revealed that students' prior KEC with biomolecules and molecular representation varied in a cluster distinct from to their KEC for experimental biology and both clusters separately from their KEC for plant physiology. Many instructors choose to focus a course more on biology content learning that is easier to test thank knowledge of experiments.

Findings derived from the rigorous methods indicate that because perceptions of knowledge, experience and confidence for 'Experimental Design' category started higher, students may have felt they were making more progress with their learning in the other categories. The network structure diagram is useful to hypothesize strategic next steps for modifying instructional activities and the design of potential future assessments.

Designing a workshop to assess biology students' learning about experimental design: Drawing from research, a workshop was designed to introduce faculty participants to two assessments: the Participant Perception Inventory (PPI) (Appendix M) and a Rubric for Experimental Design (RED) (Chapter 2, Table 2.2). Participants first examined raw student data from the beginning and mid-way through a CURE. Analysis templates and handouts helped them evaluate students' KEC with experimental biology. The PPI allows a quick measurement that can be used to guide instructional strategies when there is still time to make changes before the end of a course. However, self-

reported perceptions can be flawed because students under- or overestimate their knowledge. Thus, in a second phase, this workshop introduced participants to a direct measure of ability to design experiments. Student data was evaluated with a Rubric for Experimental Design (RED) to indicate knowledge of, and difficulties with, experimental design. Participants practiced using the RED in a third phase of the workshop. In small groups, they decided, based on the data, what instructional experiences to provide next. At the end of this workshop, participants were able to (a) design a PPI for their own target learning outcomes, (b) diagnose students' experimental design knowledge using RED, and (c) consider how to address problems based on two complimentary measures of experimental design learning.

<div align="center">5.4 <u>Scope of Future Research</u></div>

While this dissertation makes considerable strides towards an in-depth knowledge of undergraduate student difficulties in experimental design, it also sets a foundation for potential future research in this nascent yet exciting stream of research. Some broad avenues of future research are highlighted below.

Effect of alternating cover stories. An interesting research direction to extend the 'Neuron Assessment' study (Chapters 3 and 4) can be the application of alternative assessment cover stories, i.e. different versions of the same subject matter used in an assessment to examine variances, if any, in students' experimental approaches. Previous research (Tshirgi, 1980; Schauble 1991) suggests that students use variable hypothesis-testing approaches depending upon the cover story used to direct an assessment task. But it also could be that by providing content with a background story and appropriate visuals, domain general skills might be identified. On the other hand, some cover stories may point toward a domain-general skills that some students may not yet have developed.

Comparison of traditional vs. reformed labs. We offer the RED as a research tool that can be used to measure experiential learning in lab-based courses. A future research direction may be to apply the RED to cross-compare knowledge of student-participants in two variable lab settings to examine if learning about particular experimental areas is

more effective in a certain kind of lab setting. For example, it will be interesting to analyze whether "reformed" labs show comparable results to traditional labs, or whether traditional labs reveal students with better experimental design knowledge.

Designing a skill-based lecture or lab module. In the study with the 'Neuron Assessment', all four participants suggested a need for increased practice with experimental design exercises. Physics educator, Joe Redish (http://umdperg.pbworks.com/w/page/10511170/121-122%20Reformed%20Labs) suggests a need to reform labs so that when students design or conduct experiments they are probed with questions like, *"What are you doing?"*, *"Why are you doing it?"* and *"If you succeed how will you get the answer to the question you are investigating?" (*With this perspective and moving forward as a future practitioner, my goal is to design teaching modules where students are trained with the experimental abilities that equip them to design their own experiments to pursue personally relevant questions.

According to the *Vision and Change* (AAAS, 2011) report, core competencies for disciplinary practices include formal practices of observation, experimentation, and hypothesis testing; applying quantitative analysis and mathematical reasoning; and using modeling and simulation to focus on the study of complex systems. Therefore it is of current relevance that undergraduate students are trained to learn about the experimental research (AAAS, 2011; Brickman et al., 2012; Hiebert, 2007; Hoskins et al., 2007; Hoskins & Stevens, 2009; Hoskins et al., 2011; Pellegrino & Hilton, 2012; PCAST, 2012; Ruiz-Primo et al., 2010; Singer et al., 2012; Wei & Woodin, 2011). This dissertation makes important contributions to the area of biology education research by establishing critical findings about student experimental design knowledge and difficulties. Further, this dissertation also investigates sources of these difficulties in order to identify specific concepts that students find problematic. With this information, useful remediation strategies can be planned. For instance, specific learning objectives can be designed according to areas that need specific attention followed by specific diagnosis (using existent or original diagnostic assessments) about whether students continue to have trouble with those areas.

Gaining appropriate knowledge about experimental research is vital for students to understand biology ranging from introductory to advanced level undergraduate courses and also provides a competitive edge for future employment in graduate school or other scientific careers. Thus, findings from this dissertation can be used to promote experimental knowledge at the undergraduate level and further open up several new avenues to be explored to progress student understanding of the experimental basis of biological phenomena.

References

AAMC-HHMI (2009) SFFP (Scientific Foundations for Future Physicians) Competences Report.

American Association for the Advancement of Science (AAAS) (2011) *Vision and Change in Undergraduate Biology Education*. AAAS, Washington, DC. http://visionandchange.org/finalreport

Brickman P, Gormally C, Francom G, Jardeleza SE, Schutte VGW, Jordan C, & Kanizay, L (2012). Media-Savvy Scientific Literacy: Developing Critical Evaluation Skills by Investigating Scientific Claims. The American Biology Teacher *74*, 374–379.

Clase KL, Gundlach, E, Pelaez NJ (2010). Calibrated Peer Review for Computer-Assisted Learning of Biological Research Competencies. Biochemistry and Molecular Biology Education *38*, 290-295.

Dasgupta, AP, Anderson TR, Pelaez, N (2014). Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties. CBE-Life Sciences Education *13*, 265-284.

Hiebert SM (2007). Teaching simple experimental design to undergraduates: do your students understand the basics? Advances in Physiology Education *31*, 82–92.

Hoskins S, Stevens L, Nehm R (2007). Selective Use of Primary Literature Transforms the Classroom into a Virtual Laboratory. Genetics *176*, 1381-1389.

Hoskins SG, Stevens LM (2009) Learning our L.I.M.I.T.S. Less is more in teaching science. Advan. Physiol. Edu. *33*, 17-20.

Hoskins SG, Lopatto D, Stevens LS (2011). The C.R.E.A.T.E. Approach to Primary Literature Shifts Undergraduates' Self-Assessed Ability to Read and Analyze Journal Articles, Attitudes about Science, and Epistemological Beliefs CBE Life Science Education *10*, 368-378.

National Research Council (2007). Rising above the gathering storm: energizing and employing America for a brighter economic future. National Academies Press, Washington, DC.

Pellegrino J, Hilton ML (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Washington DC: National Research Council.

President's Council of Advisors on Science and Technology (PCAST) (2012). Transformation and opportunity: the future of the U.S. research enterprise. Washington, DC. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_future_research_enterprise_20121130.pdf

Ruiz-Prim MA, Li M, Tsai SP, Schneider J (2010). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. Journal of Research in Science Teaching *47*, 583–608.

Schauble L (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. Journal of experimental child psychology *49*, 31-57.

Schönborn KJ, Anderson TR (2009). A model of factors determining students' ability to interpret external representations in biochemistry. International Journal of Science Education *31*, 193-232.

Singer SR, Nielsen NR, Schweingruber HA (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. Washington DC: National Research Council.

Tschirgi JE (1980). Sensible reasoning: A hypothesis about hypotheses. Child development, 1-10.

Wei C, Woodin T (2011). Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE Life Science Education *10*, 123–31.

APPENDICES

Appendix A: The 'Shrimp Assessment'

Background Information

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

1. List the treatments that the biologist plans to use in this experiment.

The three different growth-enhancing nutrients (A, B, and C) and two different salinity levels (low and high) yield a total of 3*2 = 6 different treatment combinations for this experiment, so each can be replicated.

Treatment	Salinity	Nutrient
1	Low	A
2	High	A
3	Low	B
4	High	B
5	Low	C
6	High	C

2. Using the treatments listed in part (a), describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks.

Since 10 tiger shrimps have already been randomly placed into each of 12 similar tanks in a controlled environment, we must randomly assign the treatment combinations to the tanks. Each treatment combination will be randomly assigned to 2 of the 12 tanks. One way to do this is to generate a random number for each tank. The treatment combinations are then assigned by sorting the random numbers from smallest to largest.

3. Give one statistical advantage to having only tiger shrimps in the experiment. Explain why this is an advantage.

Using only tiger shrimp will reduce a source of variation in the experimental units, the tanks of shrimp in this experiment. By eliminating this possible source of variation, type of shrimp, we are better able to isolate the variability due to the factors of interest to

us (nutrient and salinity level). This will make it easier to identify any treatment effects that may be present.

4. Give one statistical disadvantage to having only tiger shrimps in the experiment. Explain why this is a disadvantage.

Using only tiger shrimp will limit the scope of inference for the biologist. Ideally, the biologist would like to identify the treatment combination that leads to the most growth for all shrimp. However, the biologist will only be able to identify the best treatment combination for tiger shrimp because other types of shrimp may respond differently to the treatments.

Appendix B: The 'Drug Assessment'

Background
The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing.

Directions
As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

1. Using complete sentences state the hypothesis to be tested.
Alamain will be successful in lowering the blood pressure in human subjects with high blood pressure levels.

2. Since there are several contributing factors that can affect blood pressure levels, list five factors that will be constant between the experimental and control groups.
Age, smoker or non-smoker, sex, present blood pressure, diet, stress, amount of daily exercise, percent body fat, weight, family history, daily or weekly alcohol consumption, cholesterol level, etc.

3. Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study.
The categories would have to be chosen to match the people in the two different groups as closely as possible to the factors listed in Question #2.

4. Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group.
I would divide up the participants randomly in the control and experimental groups.

5. Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken.

I would check their blood pressure and heart rates at least once a day, once a week, etc. and measure any side effects between the two groups.

6. Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans.

The drug lowered the blood pressure in the experimental group with no harmful side effects.

Appendix C: The 'Bird Assessment'

The College Board (2009) AP® Statistics Free-Response Form B. Question 4 Page 8.
Assessment: [Online
http://apcentral.collegeboard.com/apc/public/repository/ap09_frq_statistics_formb.pdf]
    Scoring Guidelines:
http://apcentral.collegeboard.com/apc/public/repository/ap09_statistics_form_b_sgs.pdf
    (Used with permission to Nancy Pelaez, npelaez@purdue.edu)

1. Birds have four types of color receptors in the eye. Most mammals have two types of receptors, although primates have three. Birds also have proportionally more nerve connections between the photoreceptors and the brain. Previous research has shown differences between male and female zebrafinches in their tendency to avoid food that has solid colors. Suggest a potential cause for this difference between male and female zebrafinches. Briefly explain.

Because birds have four types of color receptors, they are able to see different wavelengths of light than mammals that have two or three types. The four color receptors also give a broader range of light, possibly allowing the birds to see ultraviolet light. Male zebrafinches are very distinct from female zebrafinches. The males have bright patches of color on their plumage, while females are mostly one solid color. Evolution may have adapted male zebrafinches to be attracted to solid colors so they will easily find a mate. This would explain why males eat solid colored fruit. On the contrary, females may have adapted to be attracted to stripes or patterns of colors. This would explain why females avoid eating solid fruit. Because they avoid solid fruit, one could say they may also avoid other solid females making their chances of mating increase.

2. Good biological knowledge could help you become an entrepreneur. For example, a manufacturer of toxic pesticide granules plans to use a dye to color the pesticide so that birds will avoid eating it. A series of experiments will be designed to find colors or patterns that three bird species (blackbirds, zebrafinches, and geese) will avoid eating. Representative samples of birds will be captured to use in the experiments, and the response variable will be the amount of time a hungry bird will avoid eating food of a particular color or pattern. a. Previous research has shown that male birds do not avoid solid colors. However, it is possible that males might avoid colors displayed in a pattern, such as stripes. In an effort to prevent males from eating the pesticide, the following two treatments are applied to pesticide granules:
    Treatment 1: A red background with narrow blue stripes
    Treatment 2: A blue background with narrow red stripes
    To increase the power of detecting a difference in the two treatments in the analysis of the experiment, the researcher decided to block on the three species of birds (blackbirds, zebrafinches, and geese). Assuming there are 100 birds of each of the three species, explain how you would assign birds to treatments in such a block design.

Form three blocks based on the species of bird (blackbirds, starlings, and geese) carrying a equal distribution of male: female birds to accomplish the goal of blocking to create groups of homogeneous experimental units. Within each of the three blocks, carry out a completely randomized design by randomly assigning the birds within each block to one of the two treatments. Within block 1, each bird of a particular species (let's say the blackbirds) will be tagged with a unique random number using a random number generator on a calculator, statistical software, or a random number table. The random numbers will be sorted from lowest to highest. The birds with the lowest 50 numbers in the ordered list will receive treatment 1 (red background with narrow blue stripes). The birds with the highest 50 numbers will receive treatment 2 (blue background with narrow red stripes). This method of randomization should be repeated in the other two blocks.

b. What else could the researcher do to increase the power of detecting a difference in the two treatments in the analysis of the experiment? Explain how your approach would increase the power.

To increase power (other than by blocking), the researcher could increase the sample size. This reduces the standard error of the sampling distribution. With a smaller standard error, a test is more likely to be able to detect a difference in results from the two treatments, if such a difference exists.

Appendix D: Typical 'Evidence of Difficulties' Examples from RED (Table 2)

Tables SI 1- 3 include response phrases that provide evidence of difficulties that are underlined and coded with a footnote that corresponds to a row in Table 2.

Table SI 1: Typical 'evidence of difficulties' from the 'Shrimp Assessment' responses.

'Shrimp Assessment': A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).			
Student ID	1. List the treatments that the biologist plans to use in this experiment.	2. Using the treatments listed in part (a), describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks.	3. Give one statistical advantage to having only tiger shrimps in the experiment. Explain why this is an advantage.	4. Give one statistical disadvantage to having only tiger shrimps in the experiment. Explain why this is a disadvantage.
Anna (Correct)	1. A Low salinity 2. A high salinity 3. B low salinity 4. B high salinity 5. C low salinity 6. C high salinity	A randomized design would be possibly dividing the 6 treatments into each of 12 tanks, so that there are two tanks with each treatment. In order for randomization to occur it might be easiest to use dice and assign each number to its corresponding treatment number. Example: Roll dice 1+ 2; Outcome Die 1= 2 and Die 2= 4. From this you would put treatment two and four in tanks 1 and 2.	The advantage to having only tiger shrimp in the experiment is that you are only using one single species of shrimp. This leads to an advantage because there is less variability within the growth of shrimp. As a result, using only tiger shrimps reduces variance.	One statistical disadvantage to only having only tiger shrimp is that due to the fact we only used one species of shrimp we are not able to make a generalization about all shrimp. Our data only correlates to the experiment performed on tiger shrimps. Therefore we can only make an accurate analysis on this particular species of shrimp.
Beth (Difficulty)	Nutrient A with low salinity, Nutrient B with low salinity, Nutrient C with low salinity, Nutrient A with high salinity, Nutrient B with high salinity, Nutrient C with high salinity, Low salinity with no nutrient, High salinity with no nutrient.[1,2]	Assign each tank a treatment. Put 12 slips of paper numbered 1-12 in a bowl. With all the shrimp in one tank, one by one randomly assign a shrimp to a tank. Replace the 12 strips to the bowl following each 12 shrimps[3]. By doing this, the biologist is aware of which tanks contain which ingredients but the shrimp are completely randomized.[4]	The tiger shrimps act as the control group[5]. In this, a researcher can confidently expect to find a repetitive response to a given exposure in a group of genetically identical tiger shrimps.[6,7]	The researcher is only studying the effects of a given ingredient on tiger shrimps. This [doesn't] demonstrate how a given ingredient may affect another type of shrimp.[8] Ultimately it limits the depth of the study.

[1] Area of difficulty 2-f
[2] Area of difficulty 2-c
[3] Area of difficulty 4-h

Table SI 2: Typical 'evidence of difficulties' from the 'Drug Assessment' responses.

'Drug Assessment': The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing. Directions: As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

Student ID	1. Using complete sentences state the hypothesis to be tested.	2. Since there are several contributing factors that can affect blood pressure levels, list five factors that will be constant between the experimental and control groups.	3. Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study.	4. Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group.	5. Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken.	6. Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans.
Josh (Correct)	The hypothesis is that the new drug will lower the blood pressure of people with high blood pressure.	They have to be at the same range of high blood pressure, diet, exercise, eating habits, sleep habits, etc.	These factors are important because without a consistency in the individuals chosen we cannot effectively judge how the drug works based on [results for] the control group and the experimental group members.	They will be chosen at random to be part of the experimental or control group. That way they do not have an opinion on how the drug may or may not be helping them.	Blood pressure will be monitored daily and recorded. The progress of people taking the drug will determine its effectiveness.	If people [with high blood pressure], in the experimental group who take the drug consistently have decreased blood pressure, then the drug is effective.

[4] Area of difficulty 4-f
[5] Area of difficulty 1-a
[6] Area of difficulty 3-e
[7] Area of difficulty 4-a
[8] Area of difficulty 5-c

Ken (Difficulty)	We are going to bring in individuals who are willing to test a new drug, Alamain, which we know have only produced good results on animals so far. This drug will be administered to people at low dosages at first[9], and then we will record results and from there calculate the correct amount of Alamain that should be given to each person.[10]	Hemoglobin levels will remain constant as well as most proteins. The blood vessels will be relaxed and blood will flow smoothly through them because they will expand. [11,12] To lower the pressure we administer hormones that constrict the vessels at a healthy rate. Red blood cells will remain at the same constant rate and will not be affected.	Participants cannot be pregnant simply[13] because it will affect the fetus differently than the adult. People older than 35 should not test the drug[14]. These criteria need to be met and not taken lightly because health problems may arise.[15]	The younger, healthier participants will be the experimental group while the not so young will be the control. [16,17]	Experimental groups will receive a couple different dosages to see how each dose affects blood pressure[18], whereas the control will be compared to the experimental to record differences. Measurements can be taken twice daily but no more than that to start for safety precautions.	If the drug does indeed reduce blood pressure, the percentage of those who[se] blood pressure [becomes] normal will be significantly high than that control group.[19, 20]

---

[9] Area of Difficulty 2-d
[10] Area of Difficulty 2-b
[11] Area of Difficulty 2-g
[12] Area of Difficulty 1-b
[13] Area of Difficulty 1-b
[14] Area of Difficulty 1-b
[15] Area of Difficulty 4-c
[16] Area of Difficulty 1-b
[17] Area of Difficulty 4-d
[18] Area of Difficulty 2-d
[19] Area of Difficulty 3-g
[20] Area of Difficulty 5-c

Additional Examples from the 'Typical Evidence of Difficulties' list from Table 2.1

Table SI 4: Examples of additional 'typical evidence of difficulties' according to RED from the 'Shrimp Assessment'

'Shrimp Assessment': A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

Student ID	1. List the treatments that the biologist plans to use in this experiment.	2. Using the treatments listed in part (a), describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks.	3. Give one statistical advantage to having only tiger shrimps in the experiment. Explain why this is an advantage.	4. Give one statistical disadvantage to having only tiger shrimps in the experiment. Explain why this is a disadvantage.
Ariel	The three different growth-enhancing nutrients (A,B, and C) and two different salinity levels (low and high).	Measure how much the shrimps grow in each one of the tanks with the independent variables in them. One tank would be the control with no salt or nutrients[21]. <u>There would then be tanks with no salt but with nutrient A in one, B in another, and C in the last.</u>[22] Then get three more tanks, all with salt, and place nutrient A in one, B in another, and again C in the last.	Size can be compared knowing that the only factors contributing to the differences in growth are from the independent variables since all the shrimp are alike.	The experiment is limited to the just tiger shrimp. This experiment would not explain whether the nutrients would affect any other shrimp other than tiger shrimp alone.
Brett	The different growth enhancing nutrients would be tested in both high and low salinity conditions, as in A in high salinity, A in low salinity, B in high, etc. Also, there would need to be control samples, where <u>shrimp were not given the nutrients</u>[23] and are in both high and low salinity water.	Assuming the shrimp were fed in the same manner, the easiest way to compare the shrimps' growth would be by comparing their weight. Since 10 shrimp are in each tank, comparing the total shrimp weight will give a better result than comparing individual shrimp weights.	The comparisons of weight will be simpler due to all shrimp being expected to grow similarly barring any outside influences	Tiger shrimp could be unaffected by either salinity changes or the nutrients, implying a certain reaction that can't necessarily be justified

---

Manipulation of Variables. [21] For the shrimp assessment, Ariel suggests treatment groups with a growth enhancing nutrient and no salinity: *"There would be tanks with no salt but with nutrient A in one, B in another, and C in the last"* which shows an error as independent variables are haphazardly applied, in scenarios when the combined effects of two independent variables are to be tested simultaneously, in this case, combination of salt and nutrients (Table 2, Area of Difficulty 2-e).

[22] Additionally Ariel also shows a difficulty with control groups when proposing treatments, *"One tank would be the control with no salt or nutrients."* Here the error is that the control group does not provide natural behavior conditions because absence of the variable being manipulated (salt or nutrients) in the treatment group, results in conditions unsuitable for the experimental subject as the shrimp won't survive in such conditions (Table 2, Area of Difficulty 2-h).

[23] Brett proposes a control where *"...shrimp were not given the nutrients"* which is inappropriate as the experimental goal is to compare among 3 different growth enhancing nutrients and not whether nutrients are required or not. Hence, the difficulty is control group treatment conditions are inappropriate for the stated hypothesis or experiment goal (Table 2, Area of Difficulty 2-i).

Table SI 5: Examples of additional 'typical evidence of difficulties' according to RED from the 'Drug Assessment'

'Drug Assessment': The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing. Directions: As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

Student ID	1. Using complete sentences state the hypothesis to be tested.	2. Since there are several contributing factors that can affect blood pressure levels, list five factors that will be constant between the experimental and control groups.	3. Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study.	4. Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group.	5. Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken.	6. Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans.
Cara	The drug is effective on people with high blood pressure.[24]	1.Asleep or awake – usually lower when sleeping / 2.Body position - lying down, sitting or standing / 3.Activity level - from not moving to extreme exertion / 4.Smoking – increases blood pressure / 5.Caffeine – increases blood pressure[25]	If the criteria is different there will be a complete different outcome.	They have to come from same age group.	I would have all of the participants sleep for six hours and take their blood pressure before that I would restrict them from having any alcohol caffeine or tobacco product. Then give them the ALAMAIN. Take their blood pressure every hour and record it.	The blood pressure both systolic and diastolic has come down to 140 and 90 after taking the ALAMAIN.
Doug	The administration of the drug Alamain to a group of patients will cause a significant decrease in blood pressure.	Weight, height, age, ethnicity, gender.	High blood pressure may have several different root causes that require different treatments, limit the effectiveness of a treatment, or even make certain treatment side effects occur.	They would be divided randomly to avoid bias.	Blood pressure would need to be measured over the course of several months as the drug would not be immediately effective and it would need to be seen if the drug remained	The effectiveness in lowering blood pressure, the mildness of the side effects, the length of effectiveness, and how many people can

[24] *Manipulation of Variables.* Cara's hypothesis (Table SI5), *"The drug is effective on people with high blood pressure"* only carries a treatment variable in the hypothesis statement but an outcome variable is missing as this statement does not mention *"the drug lowering blood pressure"* as a specific outcome (Table 2, Area of difficulty 2-a).

[25] Cara considers irrelevant variables in her experiments by suggesting that properties like, *"Asleep or awake, body positions"* to be maintained constant across experimental groups (Area of difficulty 2-g).

Table SI 5: Examples of additional 'typical evidence of difficulties' according to RED from the 'Drug Assessment'

'Drug Assessment': The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing. Directions: As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

					constantly effective. Initial conditions would also have to be measured to compare to later to check for side effects.	be helped by this drug would be useful criteria in measuring the drug[26].
Emma	Because the drug has been proven to be effective in animals, it will be just as effective in humans.	Five factors that should be constant are age, race, medical history, weight, and diet.	In order to test this drug, participants need to be chosen carefully. Weight should be criteria because an obese person is much more likely to have high blood pressure than a person who is of average weight. Also, the diet of the participants need to be taken into special consideration because the blood pressure of someone who eats foods that are high in fat will be much higher than that of a person who eats low-fat foods.	If all the participants fit the criteria, then they can be randomly chosen to be in either group.	The blood pressure of both groups should be taken every week and the results should be compared so as to determine if there is any change in blood pressure levels.	If the results observed in the human experiment is the same, or similar, to that observed in the animal experiment, then the drug is a success. If the results are completely different, then the drug is a failure.[27]
Frieda	ALAMAIN will safely lower blood pressure in humans and have no harmful results.	Gender, age, race, heart conditions, blood pressure range	If you are going to compare two groups, the background has to be similar/same in order to eliminate other variables that could disrupt the results.	Once a certain race is determined, then random selection would be the best. Volunteers will be	Blood pressure should be measured when resting and when exercising. Then the recovering pressure can be measured. It should also be	Long term blood pressure recovery is the best method to make sure the pressure remains low

---

[26] ***Measurement of Outcome.*** Doug's hypothesis indicates the administration of the drug Alamain is supposed to be for a group of patients and not for a large population. But when asked to suggest determination of success of the drug he states, *"How many people can be helped by this drug…"* which suggests an incoherent relationship between treatment and outcome variable (Area of difficulty 3-a).

[27] As a measure to indicate success of the blood pressure drug, Emma writes, *"If the results observed in the human experiment is the same, or similar, to that observed in the animal experiment, and then the drug is a success. If the results are completely different, then the drug is a failure."* This shows an error that an outcome variable was not listed for the investigation as we don't know what the student means by results being *"similar or different"* (Area of difficulty 3-f).

Table SI 5: Examples of additional 'typical evidence of difficulties' according to RED from the 'Drug Assessment'

'Drug Assessment': The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing. Directions: As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

				asked to join the experiment.	measured every day to make sure it isn't just short term, but long term recovery.	forever and not just when initially taken.[28]
Gage	The clinical trials of this drug will be successful by lowering patient's blood pressure.[29]	The person's blood type, cholesterol levels, genetic information, body type, and pre-existing medical conditions.	The new drug may not work on people with a certain blood type or pre-existing condition that may already alter the blood pressure. The cholesterol may inhibit the workings of the drug. Body type may play a role in how the drug is dispersed within the body. Genetic information may make someone naturally immune to the drug.	Certain blood tests would be run. A thorough medical background check would also be necessary to look for any genetic problems or pre-existing conditions that may negatively affect the drug.	Regular testing of blood coagulation would be taken to measure if the blood gets thinner or thicker.[30] I would also take regular measurements of cholesterol levels and blood pressure.	We would have to prove that patients on Alamain had regular and consistent drops in their blood pressure with minimal to no side effects. This would prove that the drug works in the human body.
Harry	ALAMAIN can lower the blood pressure of humans.	The diet menu, the time and kinds of sporting, the living habits and the age, gender and species of humans of the experimental and control group.	Because in this experiment we just want to check the effect of ALAMAIN on the blood pressure of humans, but the factors listed in Question 2 can also affect experiment results.	We have one control group and one experiment group. Just divide all the participants into these two groups randomly.	Measurement: the blood pressure of participants. / How often: three times a day: in the morning after breakfast, at the noon after lunch and at night before sleep.	Whether others can redo this experiment with other participants later and get the same result.[31]

---

[28] The stated outcome by Frieda is not measurable (Area of difficulty 3-d) as it suggests, *"Long term blood pressure recovery is the best method to make sure the pressure remains low forever and not just when initially taken."* Measuring blood pressure for a certain fixed time period is a feasible measure but *"remaining low forever"* is not when deciding success of developed drug.

[29] Gage shows an error in this area because according to his hypothesis, *"The clinical trials of this drug will be successful by lowering patient's blood pressure"* the treatment and outcome variables are reversed (Area of difficulty 3-b) as this statement implies *" success of the drug"* being the outcome variable while *"lowering blood pressure"* as the treatment or independent variable. It would be accurate if administration of drug was considered as the treatment variable and lowering of blood pressure as outcome variable.

[30] Gage also considers measurement of outcome variables *("blood coagulation testing")* that are irrelevant with his hypothesis (Area of difficulty 3-c).

[31] *Accounting for variability.* Harry suggests, *"Whether others can redo this experiment with other participants later and get the same result"* as a measure for indicating drug success which shows a problem with replication because he considers replication as repeating the entire experiment at another time with another group of experimental subjects (Table 2, Area of difficulty 4-g).

Table SI 5: Examples of additional 'typical evidence of difficulties' according to RED from the 'Drug Assessment'

'Drug Assessment': The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing. Directions: As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

Ina	The drug will be administered to a large group and variation of human subjects and will yield results that will show lower blood pressure levels.	Nutrition, stress, fitness, medication, and smoking will all be constant in the experimental group.	Nutrition is important to make sure an unhealthy or healthy food intake does not throw off results yielded from testing the drug. / Stress greatly increases blood pressure, this needs to be kept constant in all subjects to allow room to make the same difference. / Fitness should be similar throughout the test subjects in order to have similar beginning footing and to give no subject an advantage. / Medications should be kept constant and no participant can be given anything additional to avoid some medication making an unexpected change. / Smoking status needs to be similar to avoid giving anyone a disadvantage.	The control group will be comprised of all identical types of people will similar body types and lifestyles. The experimental group can have more of a variation and will be administered with the drug.[32]	Blood pressures will be regulated before each dose of Alamain (possibly once a day) and the data will be compiled and analyzed at the end of the study.	The criteria to determine success or failure will be whether the drug causes a significant negative change in blood pressure of the human test subject.

---

[32] Ina shows errors in explaining participant selection: *"The control group will be comprised of all identical types of people with similar body types and lifestyles. The experimental group can have more of a variation and will be administered with the drug."* This is an error because criteria for selecting experimental subjects for treatment vs. control group are biased (body types identical vs. variable) (Table 2, Area of difficulty 4-b). Other problems with variability are found from Ina's suggestion, *"control group will be comprised of all identical types of people"* which indicates flawed understanding of natural variability within a sample of experimental subjects (Area of difficulty 4-a). She also doesn't consider random assignment of control and experimental group participants (Area of difficulty 4-e).

Table SI 6: Examples of additional 'typical evidence of difficulties' according to RED from the 'Bird Assessment'

'Bird Assessment': Birds have four types of color receptors in the eye. Most mammals have two types of receptors, although primates have three. Birds also have proportionally more nerve connections between the photoreceptors and the brain. Previous research has shown differences between male and female zebra finches in their tendency to avoid food that has solid colors. A manufacturer of toxic pesticide granules plans to use a dye to color the pesticide so that birds will avoid eating it. A series of experiments will be designed to find colors or patterns that three bird species (blackbirds, zebra finches, and geese) will avoid eating. Representative samples of birds will be captured to use in the experiments, and the response variable will be the amount of time a hungry bird will avoid eating food of a particular color or pattern. Previous research has shown that male birds do not avoid solid colors. However, it is possible that males might avoid colors displayed in a pattern, such as stripes. In an effort to prevent males from eating the pesticide, the following two treatments are applied to pesticide granules: Treatment 1: A red background with narrow blue stripes; Treatment 2: A blue background with narrow red stripes.

Student ID	1. Suggest a potential cause for the difference between male and female zebra finches. Briefly explain.	2. a. To increase the power of detecting a difference in the two treatments in the analysis of the experiment, the researcher decided to block on the three species of birds (blackbirds, zebrafinches, and geese). Assuming there are 100 birds of each of the three species, explain how you would assign birds to treatments in such a block design.	b. What else could the researcher do to increase the power of detecting a difference in the two treatments in the analysis of the experiment? Explain how your approach would increase the power.
Jack	A potential cause for male and female Zebra Finches difference's in avoiding food that has solid colors could be the result of females needing a certain protein that are found in certain solid or non-solid foods. This may be important in the development of healthy chicks. The males may eat certain solid or non-solid foods in order for the coloration on their feathers to show up brighter. For example, Flamingos eat shrimp that cause the pink coloration of their feathers. It could also hold true for the male Zebra Finch, in order to help attract a mate.	For treatment one, the researcher should test fifty male birds of each species to understand which species of male will avoid a red background with narrow blue stripes. Treatment two will have the remaining fifty male birds of each species in order to understand which species avoids a blue background with narrow red strips. Each species will be tested separately of each other.	The researcher could <u>test different size objects and shapes with either a red background with narrow blue stripes or a blue background</u> [33]with narrow red stripes. This would help the researchers in determining which granules need to be patterned if they know the size of the birds feed. The researcher can also use different colors for testing, such as orange and blue or orange and red. Testing different colors may allow the manufacturer to use more than one patterning of colors or enable them to use the cheaper color that would be used in the dye. It is also a good idea because one or none of the species of birds will avoid seeds in either treatment.

---

[33] **Measurement of outcome.** We found an example of a response by Jack elucidating a difficulty with this area because he suggests to increase the power of detecting a difference in treatments as: *"...test different size objects and shapes with either a red background with narrow blue stripes or a blue background"* This indicates Jack proposes outcome variables (like "size, shapes, variable patterning, price of color") that are irrelevant for his proposed experimental context or provided treatments ("testing how long a bird will avoid colors displayed in stripes") (Table 2, Area of difficulty 3-c).

Appendix E: Inter-rater Reliability Results

10 responses were coded for each assessment. Steps followed for inter-rater reliability exercise are:

- Detailed explanation of rubric in terms of propositional statements for each category, concepts associated with each category and corresponding errors descriptions.
- Explanation of scoring protocol.
- One example for each assessment coded together as an example.
- Raters separated and coded individually.
- Get back together and discuss coding.
- Discuss queries/areas that need clarifications, if any.
- Determine Cohen's kappa values for each area.

Cohen's *kappa* is calculated using the formula $kappa = \frac{f_0 - f_c}{N - f_c}$ where f_0 denotes the number of responses coded similarly, f_c denotes number of responses that would be expected to be coded the same way by chance alone, and N is the number of units coded by either coder (i.e., if two coders code 50 responses each, N = 50). We calculated *kappa* values for 10 responses from each assessment and compared agreement for 5 major areas. For example, table 1 represents the coding results for the 'Shrimp Assessment'.

Table SI 7: Frequency of Correct vs. Difficulty for 'Shrimp Assessment' by raters A and B

'Shrimp Assessment'		Rater B		Rater A total
		Correct	Difficulty	
Rater A	Correct	15	0	15
	Difficulty	1	31	32
	Rater B total	16	31	47

Number of areas coded as 'correct' by both raters A and B are 15 and number of areas coded as 'difficulty' by rater A but coded 'correct' by rater B is 1. Similarly coded areas by both raters are tallied in the diagonal of the table.

Frequency of areas coded similarly, f_0, was 46 (97.87% of codes). Frequency of areas expected to be coded similarly by chance, f_c is calculated using formula:

$$fc = \frac{\text{Rater A}_{\text{correct total}} * \text{Rater B}_{\text{correct total}} / \text{Grand Total} + \text{Rater A}_{\text{difficulty total}} * \text{Rater B}_{\text{difficulty total}} / \text{Grand Total}}{\text{Grant Total}}$$

Thus, fc=$\dfrac{\frac{15*16}{47}+\frac{32*31}{47}}{47}$=0.56 or 56%. This means f_c is 56% of 46 (frequency of codes coded

similarly) is 26.2. Thus inserting these values into the formula for $kappa = \dfrac{f_0 - f_c}{N - f_c} =$

$\dfrac{46-26.2}{47-26.2} = 0.952$.

Interrater reliability was established over 50 RED areas [10 (responses) x 5 (areas)] but for kappa calculations we consider only 47 because 3 areas were classified under 'lack of evidence' and we calculated *kappa* values only for areas coded as 'correct' and 'difficulty'.

Apply the same calculations, *kappa* values for the 'Drug' and the 'Bird Assessment' was found to be 0.929 and 0.896 respectively as shown below.

Table SI 8: Frequency of Correct vs. Difficulty for 'Drug Assessment' by raters A and B

'Drug Assessment'		Rater B		Rater A total
		Correct	Difficulty	
Rater A	Correct	10	0	10
	Difficulty	1	44	45
	Rater B total	11	44	55

Number of observed agreements: 54 (98.18% of the observations). Number of agreements expected by chance: 38.0 (69.09% of the observations). Kappa= 0.929.

Table SI 9: Frequency of Correct vs. Difficulty for the 'Bird Assessment' by raters A and B

'Bird Assessment'		Rater B		Rater A total
		Correct	Difficulty	
Rater A	Correct	13	1	14
	Difficulty	2	36	38
	Rater B total	15	37	52

Number of observed agreements: 49 (94.23% of the observations). Number of agreements expected by chance: 31.1 (59.76% of the observations). Kappa= 0.857

Table SI 10: Frequency of 'correct' and 'difficulty' experimental design areas as measured by three assessments pre (beginning) and post (after) semester.

Areas of Experimental Design Difficulty	'Shrimp Assessment'	Pre (spring 2010; n =40[a])	Post (spring 2009; n =40[b])	p-value[c] from Fisher's test	Interrater Agreement[d] (Cohen's *kappa*)
Variable Property of an Experimental Subject	Correct	19	31	0.019**	
	Difficulty	18	9		
Manipulation of Variables	Correct	4	17	0.008***	
	Difficulty	27	22		
Measurement of Outcome	Correct	11	24	0.114	0.90[+]
	Difficulty	9	6		
Accounting for Variability	Correct	3	11	0.040**	
	Difficulty	33	29		
Scope of Inference	Correct	2	13	0.004***	
	Difficulty	32	26		

Areas of Experimental Design Difficulty	'Drug Assessment'	Pre (spring 2012; n =31[a])	Post (spring 2011; n =40[b])	p-value[c] from fisher's test	Interrater Agreement[d] (Cohen's *kappa*)
Variable Property of an Experimental Subject	Correct	13	31	0.003***	
	Difficulty	18	9		
Manipulation of Variables	Correct	4	13	0.092*	
	Difficulty	26	27		
Measurement of Outcome	Correct	8	25	0.007***	0.94[+]
	Difficulty	21	15		
Accounting for Variability	Correct	8	18	0.134	
	Difficulty	22	21		
Scope of Inference	Correct	2	9	0.096*	
	Difficulty	28	29		

Table SI 10 continued

Areas of Experimental Design Difficulty	'Bird Assessment'	Pre (spring 2011; n =40[a])	Post (spring 2010; n =40[b])	p-value[c] from fisher's test	Interrater Agreement[d] (Cohen's *kappa*)
Variable Property of An Experimental Subject	Correct	12	16	0.482	
	Difficulty	27	24		
Manipulation of Variables	Correct	4	14	0.015**	
	Difficulty	35	26		
Measurement of Outcome	Correct	9	16	0.025**	0.86+
	Difficulty	18	8		
Accounting for Variability	Correct	4	7	0.516	
	Difficulty	34	31		
Scope of Inference	Correct	2	6	0.264	
	Difficulty	33	32		

[ab] Categories where frequency for correct and difficulty is less than the total *n* indicates that remaining responses were classified under 'Lack of Evidence' in those cases.
[c]*p<0.01 = ***; p<0.05**; p<0.1 =*
[d]According to Landis and Koch (1977) a *kappa* value >0.70+ indicates a high degree of interrater agreement .

Table SI 11: Pre and post % <u>differences</u> in 'correct', 'difficulty' and 'lack of evidence' for five areas of experimental design knowledge

'Shrimp Assessment'	Variable property of an experimental subject (%)	Manipulation of Variables (%)	Measurement of Outcome (%)	Accounting for Variability (%)	Scope of Inference (%)
Correct	29.5	32.5	32.5	20	27.5
LOE	-8	-20	-25	-10	-12.5
Difficulty	-22.5	-12.5	-7.5	-10	-15

'Drug Assessment'	Variable property of an experimental subject (%)	Manipulation of Variables (%)	Measurement of Outcome (%)	Accounting for Variability (%)	Scope of Inference (%)
Correct	35.56	19.60	36.69	19.19	16.05
LOE	0.00	-3.23	-6.45	-0.73	1.77
Difficulty	-35.56	-16.37	-30.24	-18.47	-17.82

'Bird Assessment'	Variable property of an experimental subject (%)	Manipulation of Variables (%)	Measurement of Outcome (%)	Accounting for Variability (%)	Scope of Inference (%)
Correct	10	25	17.5	7.5	10
LOE	-2.5	-2.5	7.5	0	-7.5
Difficulty	-7.5	-22.5	-25	-7.5	-2.5

Appendix F: Glossary of Terms (in alphabetical order)

1. **Control:** An experimental baseline against which an effect of the treatment conditions may be compared (Holmes, Moody & Dine, 2011).

2. **Control group**: the "untreated" group with which an experimental group (or treatment group) is contrasted. It consists of units of study that did not receive the treatment whose effect is under investigation (Gill & Walsh, 2010).

3. **Correlation relationship:** Two variables are said to be correlated if an observed change in the level of one variable is accompanied by a change in the level of another variable. The change may be in the same direction (positive correlation) or in the opposite direction (negative correlation). Note that correlation does not imply causality. It is possible for two variables to be associated with each other without one of them causing the observed behavior in the other. When this is the case it is usually because there is a third (possibly unknown) causal factor (NIST/SEMATECH, 2003).

4. **Cause and effect relationship:** There is a causal and effect relationship between two variables if a change in the level of one variable (independent variable) causes an effect in the other variable (dependent variable). To establish a cause and effect relationship, one must gather the data by experimental means, controlling unrelated variables which might confound the results. Having gathered the data in this fashion, if one can establish that the experimentally manipulated variable is correlated with the dependent variable, then one should be (somewhat) comfortable in making a causal inference. That is, when the data have been gathered by experimental means and confounds have been eliminated, correlation does imply causation (NIST/SEMATECH, 2003; Wuensch, 2001).

5. **Factors:** the specific treatments or experimental conditions (the independent variables) (Dasgupta et al., 2013).

6. **Hypothesis**: A testable statement that carries a predicted association between a treatment and outcome variable. An investigator designs an experiment to test the hypothesis, and the experimental results are used to evaluate the hypothesis for confirmation or refutation (Ruxton & Colegrave, 2006).

7. **Outcome (dependent) variable:** A factor under investigation where it is reasonable to aruge that there may be a relationship with an independent variable. The dependant variable is measurable in terms of units. (Holmes, Moody & Dine, 2011).

8. **Outside/unrelated/control/confounding variables:** Any factors (s) that may influence your observations/experiment but is not the factor you are investigating. (Holmes, Moody & Dine, 2011).

9. **Population:** All individuals of a defined group appropriate for collecting information for a particular investigation goal (Dasgupta et al., 2013).

10. **Random (representative) sample:** A sample where all experimental subjects from a target demographic have an equal chance of being selected in the control or treatment group. An appropriate representative sample size is one that averages out any variations not controlled for in the experimental design (The College Board, 2006).

11. **Randomization:** A random sample is selected from a target population; units are then assigned to different treatment groups (Ramsey & Schafer, 2002).

12. **Replication:** Replication is performed to assess natural variability, by repeating the same manipulations to several experimental subjects (or units carrying multiple subjects), as appropriate under the same treatment conditions (Quinn & Keough, 2002).

13. **Sample:** A random (smaller) group of representative individuals selected from the population, from which data is collected and conclusions are drawn about the population (Dasgupta et al., 2013).

14. **Subject:** The individuals to whom the specific variable treatment or experimental condition is applied. Each experimental subject carries a variable property (Dasgupta et al., 2013).

15. **Treatment (independent) variable:** The factor (s) in your experiment whose effect you are examining (Holmes, Moody & Dine, 2011).

16. **Treatment group:** A group of experimental subjects or units that are exposed to experimental conditions varying in a specific way (Dasgupta et al., 2014).

17. **Unit:** The group of individuals to which the specific variable treatment or experimental condition is applied (Dasgupta et al., 2014).

18. **Variable:** A certain property of an experimental subject that can be measured and that has more than one condition (Dasgupta et al., 2014).

19. **Variation:** when observations within your data set do not all have the same value (Holmes, Moody & Dine, 2011).

20. **Variability:** sources of variability in the experimental design of biological study are often divided into two categories: biological variability (variability due to subjects, organisms, and biological samples) and technical variability (variability due measurement, instrumentation, and sample preparation) (Box et al. 2005; Cox and Reid 2000).

Appendix G: Modified Glossary of Terms

(Modified based on 'Neuron Assessment'; in alphabetical order)

Note: Underlines indicate modifications to glossary from Dasgupta et al., 2014

1. **Control:** An experimental baseline against which an effect of the treatment conditions may be compared (Holmes, Moody & Dine, 2011). <u>The control variable is represented on the x-axis in comparison to the treatment group in a graph or as a comparison set of data in the graph</u>.

2. **Control group:** A control group of experimental subjects or units, for comparison purposes, measures natural behavior under a normal condition instead of exposing them to experimental treatment conditions. Parameters other than the treatment variables are identical for both the treatment and control conditions. (Gill and Walsh, 2010; Holmes, Moody and Dine, 2011).

3. **Correlation relationship:** Two variables are said to be correlated if an observed change in the level of one variable is accompanied by a change in the level of another variable. The change may be in the same direction (positive correlation) or in the opposite direction (negative correlation). Note that correlation does not imply causality. It is possible for two variables to be associated with each other without one of them causing the observed behavior in the other. When this is the case it is usually because there is a third (possibly unknown) causal factor (NIST/SEMATECH, 2003)

4. **Cause and effect relationship:** There is a causal and effect relationship between two variables if a change in the level of one variable (independent variable) causes an effect in the other variable (dependent variable). To establish a cause and effect relationship, one must gather the data by experimental means, controlling unrelated variables which might confound the results. Having gathered the data in this fashion, if one can establish that the experimentally manipulated variable is correlated with the dependent variable, then one should be (somewhat) comfortable in making a causal inference. That is, when the data have been gathered by experimental means and confounds have been eliminated, correlation does imply causation (NIST/SEMATECH, 2003; Wuensch, 2001). <u>The causal relationship would be coherently interpreted from a graphical representation if one is included</u>.

5. **Factors:** the specific treatments or experimental conditions (the independent variables) (Dasgupta et al., 2013). <u>These are identified in a key, the symbols and figure legend.</u>

6. **Hypothesis**: A testable statement that carries a predicted association between a treatment and outcome variable. An investigator designs an experiment to test the

hypothesis, and the experimental results are used to evaluate the hypothesis for confirmation or refutation (Ruxton & Colegrave, 2006).
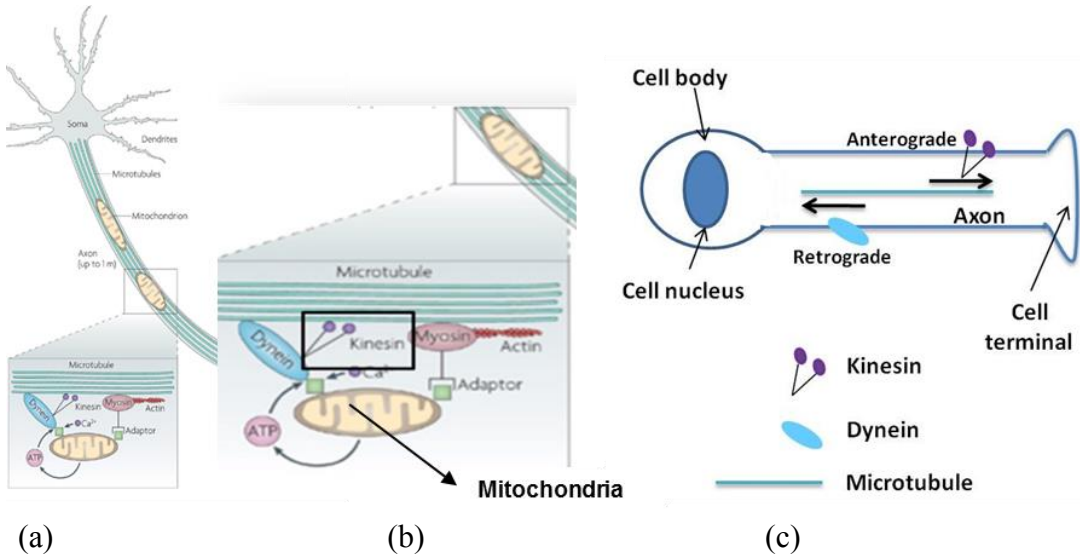
7. **Outcome (dependent) variable:** A factor under investigation where it is reasonable to argue that there may be a relationship with an independent variable. The dependent variable is measurable in terms of units. (Holmes, Moody & Dine, 2011). In a graph, appropriate outcome variables would be on the y axis.

8. **Outside/unrelated/control/confounding variables:** Any factors (s) that may influence your observations/experiment but is not the factor you are investigating. (Holmes, Moody & Dine, 2011).

9. **Population:** All individuals of a defined group appropriate for collecting information for a particular investigation goal (Dasgupta et al., 2013).

10. **Random (representative) sample:** A sample where all experimental subjects from a target demographic have an equal chance of being selected in the control or treatment group.

11. **Randomization:** A random sample is selected from a target population; units are then assigned to different treatment groups (Ramsey & Schafer, 2002).

12. **Replication:** Replication is performed to assess natural variability, by repeating the same manipulations to several experimental subjects (or units carrying multiple subjects), as appropriate under the same treatment conditions (Quinn & Keough, 2002).

13. **Sample:** A random (smaller) group of representative individuals selected from the population, from which data is collected and conclusions are drawn about the population (Dasgupta et al., 2013).

14. **Sample size:** An appropriate representative sample size is one that averages out any variations not controlled for in the experimental design (The College Board, 2006).

15. **Scope of inference**: Recognizing the extent and limit of inferences that can be made from a small characteristic sample of experimental subjects or units to a wider target population and knowing to what extent findings at the experimental subject level can be generalized.

16. **Subject:** The individuals to whom the specific variable treatment or experimental condition is applied. Each experimental subject carries a variable property (Dasgupta et al., 2013). Subjects are identified in the legend of a graph.

17. **Treatment (independent) variable:** The factor (s) in your experiment whose effect you are examining (Holmes, Moody & Dine, 2011). Treatment variables are presented as column in a table and alongside control groups on the x-axis of a graph.

18. **Treatment group:** A group of experimental subjects or units that are exposed to experimental conditions varying in a specific way (Dasgupta et al., 2014).

19. **Unit:** The group of individuals to which the specific variable treatment or experimental condition is applied (Dasgupta et al., 2013)

20. **Variable:** A certain property of an experimental subject that can be measured and that has more than one condition (Dasgupta et al., 2013).

21. **Variation:** when observations within your data set do not all have the same value (Holmes, Moody & Dine, 2011). Variations in data can be accounted for by using measures from strategies like randomization and replication.

22. **Variability:** sources of variability in the experimental design of biological study are often divided into two categories: biological variability (variability due to subjects, organisms, and biological samples) and technical variability (variability due measurement, instrumentation, and sample preparation) (Box et al. 2005; Cox and Reid 2000). <u>On a graph representing averages of experimental outcome findings, errors bars would represent variability of results from replication of treatments</u>.

Appendix H: 'Neuron Assessment' Answer

Note: This is not the only way to get a correct answer

Figures

(a)   (b)   (c)

Background

Mitochondria are one of the several organelles that get transported across the axon of a nerve *(Refer figure above)*. They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions

Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder caused by impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons.

The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;

Compound D: inhibits dynein;

An Image software: measures mitochondrial movement in neurons.

- **How do you think a 'hypothesis' relates to an experiment?**

A hypothesis is testable outcome of an experiment and defines the relationship between independent (treatment) and dependent (outcome) variables within an experiment.

1. **Describe what you see in the three diagrams above. Please tell us in detail what you think about it.**

In the left most Figure, I see the figure of an axon and mitochondria present within it. The figure in the middle is a magnified version of the mitochondria attached to microtubules via several motor proteins. The figure on the extreme right shows kinesin and dynein motor proteins that are involved in movement in the anterograde and retrograde direction respectively. The three figures together show the mechanism of movement of mitochondria along an axon with the help of motor proteins like kinesin and dynein.

2. **What could be a potential hypothesis for your experiment?**

Inhibition of kinesin and/or dynein will stop movement of mitochondria along the axon.

3. **Which factors will you vary and which will you keep the same in your study? Why?**

I would start off varying kinesin activity using compound K and observe its effect on mitochondrial movement in the anterograde direction towards the cell/axon terminal. Next I would wash off compound K to restore kinesin activity and vary dynein activity by using compound D to inhibit it. Then, I would measure movement of mitochondria in the retrograde direction. I can also use compound K and D together to see if movement of mitochondria is completely stopped across the neuron. The neuron source and other variables like calcium concentration, ATP molecules should be maintained as close as possible to reduce the effect of any confounding variables.

4. **How will you assign subjects to groups for your experimental study? Explain.**

I will ensure that I select neuronal cell cultures from pool of subjects that are representative of a larger population that the study will be applicable to. I will assign cell cultures to an **experimental** and a **control** group in my study. Cultures will be assigned to either of the groups using random sampling. The control groups cell cultures will not be treated neither compound K nor D. The experimental group will consist of cell cultures that will be treated with compound K and/or compound D.

5. Do you think you can establish a cause-and-effect relationship between the treatment and a response variable in this experiment? Justify your answer.

Yes I think a cause and effect relationship can be established between inhibition of kinesin or dynein using compound K or D (treatment) and effect of movement of mitochondria (response) if: Inhibition of kinesin using compound K stops anterograde movement; inhibition of compound D using dynein stops retrograde movement; using compound K and D in combination will complete stop or allow minimal mitochondrial movement across neurons.

6. How would you present the results of your experiment?

I would present the results with the help of a graph that will include mean mitochondrial movements towards the cell terminal (after using Compound K to inhibit Kinesin) and towards the cell body (after using Compound D to inhibit Dynein). I will also have errors bars for bars on my graph to represent mitochondrial movement variations as a result of replication of treatments.

7. What results do you expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of your experiment.

I would expect to see inhibition of kinesin result in a slowing of anterograde movement while inhibition of dynein would result in a slowing of retrograde movement. I also expect the combination of the two inhibitors would prevent any mitochondrial movement. These expectations would be validated through the use of microscopy and a digital measurement of the distance traveled.

8. How will you improve the validity of your experiment?

The findings of this experiment can be improved by repeating /replicating treatments. Also, conducting the experimental study on sample of subjects that are representative of a larger population of human subjects increases the experiment reliability.

9. **What do you think this diagram is not showing? Explain your answer.**

The diagram fails to show how the motor appears during each of the two directions of motion. But together with the figures and the background, the question has all the details necessary to answer the questions given.

10. Is there anything about this question that you don't understand or find confusing? Explain.

Not necessarily. I know you did it to simplifying the context but I believe a large body of initial work would be required to get to narrowing down to kinesin or dynein being responsible for the disorder. So in a way I like that the question makes it easy by ruling out any other possibilities because just by itself, mitochondrial transport impairment could be potentially due to a host of things.

11. Consider yourself a diagram designer. If you could change the diagrams, what would you change or how would you improve them?

The figures by themselves are OK. I know it doesn't include any measurement values because part of the question was for the students to think about that aspect. If you were to think about a classroom activity using this question, you would have the students go through the background information and perhaps sketch out plots and have that as supplement to the text.

Appendix I: Rubric for Experimental Design (RED) Including Graphical Representation Abilities

Note: Underlines indicate modifications to glossary from Dasgupta et al., 2014

Broad Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
(1) Variable property of an experimental subject	**Experimental subject or units:** The individuals to which the specific variable treatment or experimental condition is applied. An experimental subject has a variable property. A variable is a certain property of an experimental subject that can be measured and that has more than one condition.	a. An experimental subject was considered to be a variable.
		b. Groups of experimental subject were considered based on a property *that diverges* from the subjects that were the target for the stated investigation or claim to be tested.
		c. Variable property of experimental subject considered is not consistent throughout a proposed experiment.
	<u>**Graphical representation:** Experimental units or subjects are identified in a title or the legend of a graph.</u>	<u>d. The experimental subject was represented as a treatment group along the x-axis.</u>
(2) Manipulation of Variables	**Testable hypothesis:** A hypothesis is a testable statement that carries a predicted association between a treatment and outcome variable.	a. Only the treatment and/or outcome variable is present in the hypothesis statement.
		b. Hypothesis does not clearly indicate the expected outcome to be measured from a proposed experiment.
	Treatment group: A treatment group of experimental subjects or units is exposed to experimental conditions that vary in a specific way.	c. Haphazard assignment of treatments to experimental units in a manner inappropriate for the goal of an experiment.
		d. Treatment conditions proposed are unsuitable physiologically for the experimental subject or inappropriate according to the goal of an investigation.
	Combinatorial reasoning: In experimental scenarios when two or more treatment (independent) variables are present simultaneously, all combined manipulations of both together are examined to observe combinatorial effects on an outcome.	a. Independent variables are haphazardly applied, in scenarios when the combined effects of two independent variables are to be tested simultaneously.
		b. Combining treatments in scenarios where the effect of two

Note: Underlines indicate modifications to glossary from Dasgupta et al., 2014

Broad Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
		different treatments are to be determined individually.
	Controlling outside variables: The control and treatment groups are required to be matched as closely as possible to equally reduce the effect of lurking variables on both groups.	c. Variables unrelated to the research question (often showing a prior knowledge bias) are mismatched across treatment and control groups.
	Control group: A control group of experimental subjects or units, for comparison purposes, measures natural behavior under a normal condition instead of exposing them to experimental treatment conditions. Parameters other than the treatment variables are identical for both the treatment and control conditions.	d. The control group does not provide natural behavior conditions because absence of the variable being manipulated in the treatment group, results in conditions unsuitable for the experimental subject. e. Control group treatment conditions are inappropriate for the stated hypothesis or experiment goal.
		f. Experimental subjects carrying obvious differences are assigned to treatment vs. control group.
	<u>**Graphical representation:**</u> <u>Both treatment and control group are presented as a column in a table and represented side by side on the x-axis in comparison to the treatment group in a graph or as a comparison set of data in the graph</u>	<u>g. Appropriate control and/or treatment groups are not presented alongside treatment groups in tables or graphs.</u>
(3) Measurement of experimental outcome	Treatment and outcome variables should match up with proposed measurements or outcome can be categorical and/or quantitative variables treatments. -A categorical variable sorts values into distinct categories. -A quantitative or continuous variable answers a "how	a. No coherent relationship between a treatment and outcome variable is mentioned. b. The treatment and outcome variables are reversed.

Appendix I: Rubric for Experimental Design (RED) Including Graphical Representation Abilities

Note: Underlines indicate modifications to glossary from Dasgupta et al., 2014

Broad Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
	many?" type question and usually would yield quantitative responses.	
	Outcome group: The experimental subject carries a specific outcome (dependent variable) that can be observed/measured in response to the experimental conditions applied as part of the treatment.	c. Outcome variables proposed are irrelevant for the proposed experimental context provided or with the hypothesis. d. Stated outcome not measurable. e. No measure was proposed for the outcome variable. f. An outcome variable was not listed for an investigation.
		g. There is a mismatch between what the investigation claims to test and the outcome variable.
	Graphical representation: In a graph, appropriate outcome variables would be on the y axis.	h. The outcome variable is not represented on the y-axis. i. No units are represented for variable represented on the y-axis
(4) Accounting for variability	Experimental design needs to account for the variability occurring in the natural biological world. Reducing variability is essential to reduce effect of non-relevant factors in order to carefully observe effects of relevant ones.	a. Claims that a sample of experimental subjects will eliminate natural variability with those subjects.
	Selection of a random (representative) sample: A representative sample is one where all experimental subjects from a target demographic have an equal chance of being selected in the control or treatment group. An appropriate representative sample size is one that averages	b. Criteria for *selecting* experimental subjects for treatment vs. control group are biased and not uniform. c. Criteria for selecting experimental subjects for investigation are

Appendix I: Rubric for Experimental Design (RED) Including Graphical Representation Abilities

Note: Underlines indicate modifications to glossary from Dasgupta et al., 2014

Broad Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
	out any variations not controlled for in the experimental design. (NYSED, 2006)	different in a way that is not representative of the target population.
	Randomized design of an experiment: Randomizing the order in which experimental subjects or units experience treatment conditions as a way to reduce the chance of bias in the experiment.	Decisions to *assign* experimental subjects to treatment vs. control group are not random but biased for each group.
		d. Random assignment of treatments is not considered.
	Randomization can be complete or restricted. One can restrict randomization by using block design which accounts for known variability in the experiment that can't be controlled.	e. Random assignment of treatments is incomplete as they show random assignment of the experimental subjects but instead, what is needed is random assignment of treatments.
	Replication of treatments to experimental units or subjects: Replication is performed to assess natural variability, by repeating the same manipulations to several experimental subjects (or units carrying multiple subjects), as appropriate under the same treatment conditions.	f. Replication means repeating the entire experiment *at some other time* with another group of experimental subjects.
		g. No evidence of replication or suggested need to replicate as a method to access variability or to increase validity/power of an investigation.
	<u>**Graphical Representation:** On a graph representing averages of experimental outcome findings, errors bars would represent variability of results from replication of treatments.</u>	<u>h. Missing error bars on graphs representing averages of experimental outcome findings on y-axis.</u>
(5) Scope of inference of findings	**Scope of inference:** Recognizing the limit of inferences that can be made from a small characteristic sample of experimental subjects or units, to a wider target population and knowing to what extent findings at the	a. The inference from a sample is to a different target population. Usually students overestimate their findings beyond the scope of the target population.

Note: Underlines indicate modifications to glossary from Dasgupta et al., 2014

Broad Areas of Difficulty	Propositional Statements/Completely Correct Ideas	Typical Evidence of Difficulties
	experimental subject level can be generalized.	
		b. No steps are carried out to randomly select experimental subjects' representative of the target population about which claims are made.
	Cause and effect conclusions: A cause-and-effect relationship can be established as separate from a mere association between variables only when the effect of lurking variables are reduced by random assignment of treatments and matching treatment and control group conditions as closely as possible. Appropriate control groups also in comparison to the treatment group also need to be considered.	c. A causal relationship is claimed even though the data shows only association between variables. Correlation does not establish causation.
	Graphical Representation: The causal relationship would be coherently interpreted from a graphical representation if one is included.	d. A causal relationship (separate from a mere association) could not be gleaned statistically from the graph because appropriate control groups were not represented on the x-axis in comparison to the treatment group in a graph.
		e. A causal relationship could not be derived as the patterns between the treatment and outcome group were represented as different from the provided experiment background.

Appendix J: Qualitative Interview Questions (based Three Phase Seated Interview Technique or 3P SIT)

Phase 1: *Investigation of a student's <u>knowledge</u> about context (neurons and organelle movement) and experimental design <u>before</u> being exposed to the background information.*

1.1. What comes to mind when I say 'neurons'?

1.2. What comes to mind when I say 'organelle movement along neurons'?

1.3. Please draw to help me understand what you mean.

1.4. Would mitochondria perhaps be in the picture somewhere?

1.5. How do scientists know the ideas that you are telling me?

1.6. What would an experiment have involved? What would they have used?

1.7. Would they have measured something? Please explain so I know more about what you are thinking.

Phase 2: *Students' use their experimental design knowledge to design an experimental in the 'Neuron Instrument' context.*

2.1. What are your thoughts about what is represented in this figure?

2.2. Why do you think this shows organelle or mitochondrial movement in a neuron?

2.3. What are the scientist/researchers trying to do in this study?

2.4. What would an experiment have involved? What would they have used?

2.5 What would the scientists have measured?

2.6. How will you use materials to conduct your experiment step by step?

2.7 What kinds of treatments will you assign?

2.8 How would you decide on the right sample to be included in your treatment/control group in your study?

2.9. What results do you expect to get and how would you record those?

2.10. Can you please share how you would represent this experiment in a graph? List the values and units of measure in your graph.

2.11 Please explain what you draw as your graph here.

2.12 Earlier you mentioned about some treatment groups. Which of those are you representing in your graph?

Phase 3: *Students evaluate and critique the 'Neuron Instrument' and the activity, thereby allowing the researchers to gain knowledge and validate their difficulties with prior knowledge and experiments exposed in the first 2 phases.*

3.1. How would you rate the questions about experiments on a 1-10 scale and why?

3.2. Is there anything about the experiment in particular that you don't understand or find confusing?

3.3. What do you think is left out of these questions about experiments? Explain your answer.

3.4. Consider yourself a question designer or textbook author. If you could change this question in any form, what would you do to improve it, if anything?

3.5. Do you think this is a good and clear question? Give reasons for your answer.

3.6. Comment on these types of questions in general, and your feelings on interpreting them.

Appendix K: Interview transcripts

1. Interview Transcript for Expert [Eric]

Interviewer: AD; Eric: E

Phase 1

AD: Hi! Eric, I am Annwesa Dasgupta (AD). How are you doing today?

E: Good!

AD: So, thank you for being here today. Alright, I would like to briefly explain some details about this activity. You just spent some time writing ideas about what you think about experiments. Now I would like to follow up your ideas by giving you an opportunity to share some thoughts verbally via a conversation based on a few questions. Are you ready to begin?

E: Yes!

AD: So, most of my questions will be related to the written survey you just completed, to help me understand your ideas. Some instructions to get started…please think freely about the questions I ask …there are no time limitations so you are free to take as much time as you wish to respond. There is no right or wrong answer to these questions. I am simply interested in your thinking about experiments. You are free to use provided materials to draw things if that helps you to express your thoughts. However, there might be certain instances where I request to visually present your ideas just so I am sure that I understand correctly. This interview will be recorded. You may choose to withdraw your participation at any time without penalty. If you have questions or need clarification at any time during this conversation, please let me know.

AD: let's talk a little bit about neurons. What's the first thought that cross your mind when I mention "neuron"?

E: when you say "neuron", I can picture a few different morphologies of the cell and the synaptic connections between them, the neuron networks with neurons is the basis of that.

AD: So how would you visually represent these ideas?

E: Let's see, I would probably draw... *(Starts drawing Figure 3.2A)*

I would draw dendrites, an axon and I will make the axon myelinated. I am drawing a circular soma and some dendritic branches going up. I would make couple of terminals, terminal boutons and the en passant bouton. I will leave off the post synaptic boutons for the moment. Then there would be dendrites which I would see in the inferior colliculus inside the auditory thalamus. Often in textbook, the spinal motor neurons are shown as the representative neurons but they are not really representative of all kinds of neurons in

the brain with a big fat axon and sparse dendrites. That's probably not true for 90% of neurons.

AD: This is a nice visual you draw here (Figure 3.2A). Tell me little bit about what comes to your mind when I say, "organelle movement along neurons"?

E: Right! This is when the microtubules come into picture. Say a spinal motor neuron that is almost close to a meter and we need a way to get materials from the cell body down to the terminal using the tracks along the axon.

AD: How will you represent your ideas about "organelle movement" in a visual format?

E: Draws figure 3.2. (Describing Figure 3.2A-B), let's assume cargo assembles in the soma after processing through ER and Golgi to package up and ready to go. Then the cargo is sorted to microtubules and kinesin. So we have microtubules bundles going down the axon and then the kinesin heavy chain help in transporting the cargo (could be organelles) across an axon in a neuron. Kinesin is a +end directed microtubule and so it takes cargo towards the neuron terminal. Several molecules get facilitated along the axon in this manner and so something of the size of an organelle can get transported like this too.

AD: Would mitochondria perhaps be in the picture (Figure 3.2B) anywhere?

E: Mitochondria could be an organelle that would be moved along. But I am not so sure of the size and I presume if it's too large, it might take a few kinesin molecules.

AD: How did scientists find out about the ideas you show in your figures (Figure 3.2A-B)?

E: Right. In terms of the organelle movement, probably through some form of live cell imaging and a fluorescent tag to tag some mitochondrial specific protein and track the fluorescence as it moves down the axon. The axons in the study obviously should be picked from the same kind of neurons, say spinal motor neurons, to avoid confounding factors that might contaminate our findings.

AD: How would be put that in form of a visual?

E: *(Draws Figure 3.2C-D)*

So in terms of materials we will have Mitochondria and GFP is the fluorescent protein tag specific to mitochondria that's coupled to the mitochondria gene. We will assume that's how it goes into the cell. Now we have GFP-tagged mitochondria and then we have microtubules which will be attached to kinesin. Basically then we will use a fluorescent microscope to track mitochondria.

AD: So in this experiment, would they be measuring something?

E: Yes! It depends on what they want to find out. If I were to assume let's say, my goal would be track the movement of the GFP labeled mitochondria (Figure 3.2C). Specifically we start measuring right around the axon hillock where the axon branches off (center image) and let's say we have a specifically identifiable particle for each mitochondria. We can then quantify the movement of the particles along a certain segment of axon observed under the microscope. So then in terms of measurement, we can measure position going from origin to end point of the imaging field and have time (in seconds) to track the movements over time (Figure 2D). I would then assign a value to each position a mitochondria (identifiable particle) is located at a certain time and how many seconds does it take to reach a certain end point-so I will be measuring velocity in terms of quantity.

AD: Under what conditions would they made these measurements?

E: At this point hopefully we have neurons that are amenable to this procedure. So we will be using multiple neurons and then set up probably assigning sets of neurons in a randomized manner to several petri-dishes. Using the method I described, we can obtain several values for the speed of mitochondria moving towards an end point in the selected field which can be averaged eventually. I am guessing since we are only tracking movement in the neurons, a control won't be necessary at this point.

AD: Summary.

E: Our goal was to measure organelle movement within the axon. To do so, we fluorescently labeled particular organelle-mitochondria along the axon and then tracked its motion using live cell microscopy. We quantified those movements by looking at multiple sets of neurons to determine the positions of mitochondria and determined velocity and see whether there are different forms of movement.

Phase 2

Probe for surface-level reasoning

AD: Now! Here is a sheet with couple of figures that are the same figures you saw in the written survey you just completed. Along with these figures, here is another sheet with some background information. I would request you to take some time to go through these sheets. Let me know when you are ready. *[Showed the figures 1a-c to the participant...gave them some time to think about what they are seeing…and followed up with these questions below...]*

E: [After couple of minutes] I think I am ready now…

AD: Great! So first, what are your thoughts about what's represented in the three figures [Referring to Figure 3.1a-c in the 'Neuron Assessment']

E: So these are showing a neuron and focusing on the axonal transport of mitochondria. There is also a enlarged version of the microtubule motors kinesin and dynein responsible for anterograde and retrograde transport respectively.

AD: What in this figure indicates you see a neuron?

E: The dendrites and an axon are typically parts of a neuron.

AD: What indicates you see transport of mitochondria like you just mentioned?

E: The arrows within Figure 3.1c tend to indicate motion.

AD: Where have you seen anything like this before?

E: Similar things in textbooks and in my own research.

AD: What the scientists trying to do in this study?

E: In this study there are trying to test the mechanism for a particular set of neurons that have impaired mitochondrial movement.

AD: What is their goal?

E: They want to figure out how to correct the impairment to be able to apply that to repair or preventing of neurons in patients with the disorder. They already are down to the idea that a defect with either kinesin or dynein is causing the disorder.

AD: Let's imagine you are the lead scientist of a group that is supposed address the goals that would just mentioned. What specific directions would you give your team to carry out this experiment using the materials provided? Also try maybe depicting it in some form of a visual like a schematic or flowchart.

E: So we will do a position vs. time of mitochondria and looking along the axons of neurons. We will have some control neurons taken from cell culture lines that basically don't show this impairment. Then we have the impaired neuron. What we expect to see then. Let me draw this out *(Draws figure 3.2E)*.

(Describing figure 3.2E) So we have a scenario 1: kinesin impaired and scenario 2: dynein impaired. Then we will have a control (normal neurons). When nothing is added, we get baseline for anterograde and retrograde speeds. With addition of compound K, we get retrograde movement only and with compound D, we will get a anterograde movement only. This will give us an estimate of the peak antero- and retrograde speeds and what to expect when we add something. *All others details were as tabulated in Figure 2E.* This is in the case where the impairment is assumed to be a loss of function.

AD: You mention "impaired" in this figure (Figure 2E). Where are the impaired neurons coming from?

E: These neurons are derived from the cell cultures of neurons of patients/cell lines with the impairment.

AD: How will you assign the treatments in the study?

E: In an ideal world, I would be blind as to the origin of the cell-so they wouldn't know whether the representative neurons are derived from the patient population or the normal human cell line. These cells will be randomly assigned to the three treatment groups which are my three columns (Figure 3.2E). So you will have nothing added first and do a series of measurements there and then you add the inhibitor compound and look to see the change over time.

AD: What is the rationale behind randomly assign the cells as you just mentioned?

E: It is a measure to reduce bias during the experiment and also to account for variability among measures.

AD: Why do you have multiple groups (Figure 3.2E)?

E: These are two sets of outcomes based whether the kinesin or dynein is impaired. It's useful to know what your predictions about an experiment would be so you can connect it back when interpreting results.

AD: So what were your predictions?

E: For scenario 1: With kinesin impaired neurons, I would expect the addition of compound K would show any change in the movement (because the impairment and inhibitor as the same impact). But with addition of compound D, I would see no movement in both the anterograde and retrograde directions along the axon.

AD: How will decide the right sample for the control vs. kinesin impaired vs. dynein impaired treatments (Figure 3.2E)?

E: Our target is the impaired mitochondrial movement. By having a positive control we know how the movements in a normal cell looks like. We also have an idea how the normal cell looks like when we have the inhibitors.

AD: What factors that you will specifically vary or keep the same in your experiment?

E: The factors kept the same would be the imaging set up, conditions of the medium, the cell culture age, time window used to measure, effective concentrations of the inhibitors *etc*. This ensures that any external sources of variation are removed in the experiment. Variation means the differences between measurements. The things we will vary are the treatments: nothing added, compound K or compound D.

AD: Let's say you perform the experimental approaches suggest, what kind of experimental results would you expect to get? How would you represent those findings?

E: First I would look at the baseline (Figure 3.2E, column 1) which could get us relatively far to understand whether the kinesin or dynein is impaired. Let's assume for convenience that our experimental with control group cells showed that dynein is

impaired. So to represent how I reached upon that finding I would ideally draw a graph (Draws figure 3.2F).

So in a control cell from normal patients (Figure 3.2F, dashes), both anterograde and retrograde movement will take place towards the end point (100 μm). In the same kind of cell from normal patients, when compound D is added, we will notice anterograde movement only in the positive direction (dots). What we observe in the normal cells upon treatment with inhibitors can be then compared with the cells from the patients with the disease to test what we find in our study actually applies to the real patients.

So we might take a patient with the disorder, and because we know that most probably the patient has dynein impairment, when we add compound K (inhibits anterograde movement), we will see zero to no movement because both proteins are shut down- one by the disease and other by the inhibitor treatment.

The conclusion from this graph is that the dynein is impaired because in the control we see some proportion of retrograde motion but with dynein impaired we see only movement in the positive direction/anterograde movement.

In my graph, I am showing basically two groups because I focused on the different outcomes you control expect to get.

AD: How will you increase the validity of your experiment?

E: By doing that multiple times. Even though we think we have similar cells and conditions, there is going to be some variability between them and we want to determine the extent of variability.

AD: People sometimes talk about hypothesis-driven research. Your thoughts?

E: Its clearly something funding agencies prefer. It tends to drive how people frame questions. Up to a point it's useful but it's not necessarily how science was carried out a first few 100 years where it was done formally. I have some training in neuro-anatomy and it starts out more observationally and then from that you can start honing in on hypothesis but without a period of "fishing expedition", it's really hard to come up with more directive hypothesis. So one way could be you either retrospectively layout your hypothesis or have a clear starting hypothesis and are careful about your observations and let them allow you to refine your hypothesis.

Phase 3

AD: How would you rate these questions on a scale of 1-10? 10 being most comfortable and 1 being I hope I don't have to ever do this again.

E: I'd say 9 because its subject matter that I know a little bit about.

AD: Is there anything in particular about this question that you don't quite understand or find confusing?

E: Not necessarily. I know you did it to simplifying the context but I believe a large body of initial work would be required to get to narrowing down to kinesin or dynein being responsible for the disorder. So in a way I like that the question makes it easy by ruling out any other possibilities because just by itself, mitochondrial transport impairment could be potentially due to a host of things.

AD: Do you think any question about experiments is left out from what I asked you?

E: I guess there is the assumption that the experiment works in a straightforward manner. So an outcome wasn't given out. It was OK for me but for the students it would probably be not something they are used to because I don't think many come in already carrying some sort of knowledge about mitochondrial movement along neurons.

AD: if you were a diagram designer, would have drawn these pictures differently (Referring to Figure 3.1a-c in the question material)

E: The figures by themselves are quite okay. I know it doesn't include any measurement values because part of the question was for the students to think about that aspect. If you were to think about a classroom activity using this question, you would have the students go through the background information and perhaps sketch out plots and have that as supplement to the text.

AD: Do you think overall it's a good and clear question?

E: I think this is a fairly clear question. You can set up the experiment in a way that will give you some form of answer so it does lead you to derive a certain answer if you have the right ideas about designing an experiment. It leaves out a lot of aspects which is good because you can then question students about those like the things to measure and the logic/design of the experiment *etc.*

Even non experts who may be overwhelmed by some of the things here, between the figures and text they will probably do okay.

AD: What is general comment about participating in such exercises?

E: Depends on the frequency and time. I am fairly happy to participate in them. It's what I do on a regular basis.

2. Interview Transcript for Juan

AD: Interviewer; Juan (J): Student

Phase 1

AD: Hi! Juan, I am Annwesa Dasgupta (AD). How are you doing today?

J: Good!

AD: So, thank you for being here today. Alright, I would like to briefly explain some details about this activity. You just spent some time writing ideas about what you think about experiments. Now I would like to follow up your ideas by giving you an opportunity to share some thoughts verbally via a conversation based on a few questions. Are you ready to begin?

J:Yes!

AD: So, most of my questions will be related to the written survey you just completed, to help me understand your ideas. Some instructions to get started…please think freely about the questions I ask …there are no time limitations so you are free to take as much time as you wish to respond. There is no right or wrong answer to these questions. I am simply interested in your thinking about experiments. You are free to use provided materials to draw things if that helps you to express your thoughts. However, there might be certain instances where I request to visually present your ideas just so I am sure that I understand correctly. This interview will be recorded. You may choose to withdraw your participation at any time without penalty. If you have questions or need clarification at any time during this conversation, please let me know.

AD: let's talk a little bit about neurons. What's the first thought that cross your mind when I mention "neuron"?

J: like an axon and mitochondria.

AD: So then what do you think when I say "organelle movement in neurons"?

J: I know that kinesin and dynein controls the movement- as I saw in the written question. But I am not sure of what their functions were so…

AD: Before this question, what did you think of organelle movement within neurons?

J: not much-I never learned of it.

AD: Can you draw your ideas about neuron and organelle movement within it?

J: [starts drawing Figure 3A] so here's the axon. And the mitochondria goes from the cell body to the terminal which is controlled by kinesin and the other direction is controlled by dynein [Figure 3.1].

AD: So how did scientists' find out about what you depict in your figure [referring to Figure 3.1]?

J: through research and experiments.

AD: I see. So what kind of experiments would they have carried out?

J: They might have done individual experiments to find out about each part of this process. And then tried to see if one part is missing, what the effect would be on the process or how their role is necessary in the process.

AD: would they have made any measurement to figure out about the process?

J: they would be measuring the degree of necessity of a certain protein [kinesin or dynein] of the process. What is its function and if a part it needed for the body to continue functioning. If its removed what would be affected. Its specific role could be stopped or it might even stop roles of other parts too.

AD: You mention, they would have performed "individual experiments". Under what conditions would they have done these experiments?

J: they could remove kinesin and see that the mitochondria will only move one way which is probably a problem. Both the motor proteins might be necessary and their removal could lead to the disorder.

AD: would you please summarize your ideas about how scientists' would find out about the cause of a disorder with mitochondrial movement in neurons in 3-4 lines?

J: ok to summarize how scientists did their experiment, they would do individual experiments on the mitochondria, kinesin and dynein and see if they are needed. If they find that there is a problem with kinesin and/or dynein, they could manufacture genetically some substitute for the missing motor proteins and observe the effect.

Phase 2

Probe for surface-level reasoning

AD: Now! Here is a sheet with couple of figures that are the same figures you saw in the written survey you just completed. Along with these figures, here is another sheet with some background information. I would request you to take some time to go through these sheets. Let me know when you are ready. *[Showed the figures to the participant...gave them some time to think about what they are seeing…and followed up with these questions below...]*

Juan (J): [After couple of minutes] Alright! I am ready now…

AD: Great! So first, what are your thoughts about what's represented in the three figures [Referring to Figure 3.1a-c in the 'Neuron Assessment']

J: This figure shows the axon and the mitochondria movement. It represents visually what kinesin and dynein functions are [refers to Figure 3.1b]. Figure 3.1c shows kind of an enlarged version of what goes on around this part of the axon.

AD: what indicates that you see an axon in this figure?

J: I know how a neuron looks and also same for an axon. But I have studied this process.

AD: What tells you that you see mitochondria moving?

J: Figure 3.1c shows and the text supplements information about anterograde and retrograde movement towards and away from the cell body with the help of kinesin and dynein.

AD: what are scientists trying to do in this study?

J: They are trying to study a disorder and improving it and seeing if a problem with kinesin or dynein is the cause of the disorder.

AD: What is goal for this study?

J: scientists want to see if kinesin or dynein malfunction is responsible in causing the disorder.

AD: How will they do that study?

J: They will set a control with all proteins in it and… [Pause]

AD: Would it help if you were to draw this out like a flowchart or a table?

J: Ok draws Figure 3.3C.

AD: how will you use the materials provided to design the experiment you just outlined in your figure [referring to Figure 3.3C]?

J: the scientists have a goal to find out does kinesin or dynein play a role in the cause of the disease. You can use compound K on neurons that lack kinesin as group 1 and use compound D on neurons that lack dynein as group 2.

AD: Why you suggest having multiple groups in your study as you show in your figure [refer to Figure 3.3C]

J: it's not one experiment-because you can't only see one group. You need like to verify your results.

AD: Tell me bit more about that idea?

J: like each group is assessing a certain compound or lack of a protein to see if only one protein is behind the disorder or both proteins have a role in the disorder. If you remove one with the patient improve?

AD: what would the right samples be for your control and group 1 and 2?

J: if you take out the neuron and place it in some atmosphere.

AD: let's say they decide use neurons as you suggest. Is there is a certain manner in which they will assign the neurons in the experiment?

J: they will select a patient with a disorder and one without the disorder and compare them and see what the differences are. And then do the experiment with neurons from patients with the disorder and use the one without the disorder as control.

AD: Based on that, what kind of results would the scientists get?

J: I predict that both proteins are necessary but the disorder patient is going to show a problem with the proteins in comparison to a patient without the disease. Maybe the disorder is that there is no anterograde movement because the mitochondria is not moving from the cell body to the cell terminal. Or in the opposite direction.

AD: Would they be measuring anything to reach to the results you suggest?

J: They'd be measuring movement of mitochondria. And they will see if the movement changes without the protein.

AD: How would you present these results?

J: my first, like, evidence would be from the imaging software in a table. A bar graph maybe…

AD: How would you draw that bar graph?

J: let's say he found that substituting kinesin with a genetically modified version has improved the disorder-makes the movement of mitochondria more effective. Then you can say movement with the disorder was this much and one without the disorder or the substituted version was normal and more effective [Draws Figure 3.3D]

Say, the second bar shows normal movement of mitochondria and the shaded bar is representing effectiveness mitochondrial movement in a person with the disorder of impaired mitochondrial movement so I am assuming there is no as effective movement.

AD: you show "effectiveness" as your y-axis. How will you measure effectiveness?

J: It will show how smooth the mitochondria moves. I am not sure what else to measure…

AD: In your table [refer to Figure 3.3C] you mentioned 2 groups and a control. Are you representing those in your graph [Figure 3.3D]?

J: this graph is for one group.

AD: So which group would this graph be for in your opinion?

J: I am not sure. I am just showing how the disorder will improve. I am not sure which group this would be for.

AD: summary!

J: I used compound K to remove kinesin and tested if that gave rise to the disorder. I would do the same thing with dynein. I will get the results but I don't know what they would be. But according to my example [refers to Figure3.3C] when kinesin is lacking and thus, replaced with a genetically modified version of kinesin protein, the patient showed improvement in mitochondrial movement.

Phase 3

AD: How would you rate these questions on a scale of 1-10? 10 being most conformable and 1 being I hope I don't have to ever do this again.

J: I would say 5 because the questions were ok but the fact that almost everyone had to draw a visual, I didn't enjoy that.

AD: is there anything in particular about this question that you don't quite understand or find confusing?

J: [For the 'neuron assessment'] I thought that kinesin and dynein function should have been more clearly stated. If it is possible to remove them and yet not harm the patient!

AD: So from the information provided, the function of kinesin and dynein were not clear to you?

J: Well I know they are required for mitochondria to move in opposite directions in a neuron but I would like to know more about what is the problem with them that gives rise to the disorder. I would have like it to be clearer.

AD: Do you think any question about experiments is left out from what I asked you?

J: not off the top of my head

AD: if you were a diagram designer, would have drawn these pictures differently (Referring to Figure 3.1a-c in the provided question material)

J: yes! I would focus a little bit more on the two proteins and on the whole process of how the disease actually occurs in patients.

AD: What is your take answering such question in general?

J: like on an exam?

AD: Sure! But even during courses as study material?

J: Not very much.

AD: tell me why?

J: well my opinion could be anything. I could predict any kind of information but I am not sure if I can get feedback on if it's correct or wrong. I am not ok with it! I like to know the right answer!

3. Interview Transcript for Eve

AD: Interviewer; Eve (E): Student

Phase 1

AD: Hi! ES *[name hidden for confidentiality]*, I am Annwesa Dasgupta (AD). How are you doing today?

Eve (E): Good!

AD: So, thank you for being here today. Alright, I would like to briefly explain some details about this activity. You just spent some time writing ideas about what you think about experiments. Now I would like to follow up your ideas by giving you an opportunity to share some thoughts verbally via a conversation based on a few questions. Are you ready to begin?

E: Yes!

AD: So, most of my questions will be related to the written survey you just completed, to help me understand your ideas. Some instructions to get started…please think freely about the questions I ask …there are no time limitations so you are free to take as much time as you wish to respond. There is no right or wrong answer to these questions. I am simply interested in your thinking about experiments. You are free to use provided materials to draw things if that helps you to express your thoughts. However, there might be certain instances where I request to visually present your ideas just so I am sure that I understand correctly. This interview will be recorded. You may choose to withdraw your participation at any time without penalty. If you have questions or need clarification at any time during this conversation, please let me know.

AD: let's talk a little bit about neurons. What's the first thought that cross your mind when I mention "neuron"?

E: Cells in your brain that have significant movement in your thinking process and anything that occurs in your body.

AD: Building on that, what comes to mind when I say "organelles moving in a neuron"?

E: specific organelles that take part in the processes needed to get neurons acting in the way they should or to produce the information they need throughout the body.

AD: That's interesting! How would you put these ideas in a drawing?

E: *Draws Figure 3.4A*

This is what I think. The cell is the neuron. I vaguely remember what it looks like because I took psychology so I kind of know the basis but since it's a neuron, it's going

to be connected to other axons and it's going to distribute the information that going through. So there's the mitochondria and what's going on in the mitochondria determines how the transport occurs. So mitochondria is going to give off the signals needed to the axon to go the other parts of the body to do whatever it was indicated to do.

AD: You draw this visual. Tell me how do scientists know what you are telling me?

E: I would assume that they have looked at quite a few brains probably through MRIs and CAT scans to see how the axons and neurons occur. They might have actually taken neurons from the brain and looked at them in a culture and see how they interact (Figure 4B).

AD: Ok. How would you put that in a drawing?

E: *Draws Figure 3.4B*

[Explaining Figure 3.4B] So through an MRI you notice areas that light up, so you could use substances that make certain areas light up under the MRI scan. An MRI might not be the best method because it's more of an outlook on the brain overall. If you want to see up-close you can then use a microscope and then you can see the cell.

AD: Great! Would they be measuring things here?

E: well you could see how the process occurs in the cell. They could watch as it happens. So they can then determine where the two proteins are present and watch as they occur.

AD: How would you specifically describe how they would have done those experiments?

E: well to be honest, I don't understand this completely as I haven't done the research. But with the basics, they would have to do things over a period of time-various experiments to compare. They would have to take a living specimen of the cells and keep it in the environment it needs to be so it functions properly. Then would watch as it occurs and inject what they need to manipulate things in the processes they observe to see what happens if they specifically change a certain thing- and how it affects the overall transport and other things.

AD: How would summarize your ideas about this experiment you proposed to discover organelle movement in neurons, in a couple of sentences?

E: well scientists are going to need to get a hold of these cells where they think a disorder is occurring and watch it as it happens. They have to get a significant amount of samples to test as they see fit. They are going to need the control which would be people that don't have the disorder. So healthy neurons and experiment would people that carry the unhealthy neurons.

AD: You mentioned, "A significant amount of samples". Tell me a bit more about that phrase.

E: I don't really know…they have to pick a number themselves but you need to the experiment multiple times and so you would have to have a decent amount of neurons from the healthy and unhealthy patients in order to conduct the experiment to compare and make sure that the results are significantly close to each other, otherwise the experiment really wouldn't be accurate. So it's not something you can just do once and expect to understand it. Multiple trials must be done.

AD: What is the value of doing multiple trials?

E: they get you further in the experiment-because if you just don't the study one time then you don't necessarily know how it's going to work differently. Since they wanted to test both motor proteins, you are going to have to test more than one anyway. You want to see how one affects it or how the other affects it or how both affect it. You can't really do all of that in a single trial. You would multiple trails for each of those and then you need to compare the end by taking averages.

Phase 2

Probe for surface-level reasoning

AD: Now! Here is a sheet with couple of figures that are the same figures you saw in the written survey you just completed. Along with these figures, here is another sheet with some background information. I would request you to take some time to go through these sheets. Let me know when you are ready. *[Showed the figures to the participant...gave them some time to think about what they are seeing…and followed up with these questions below...]*

E: I am ready now…

AD: Great! So first, what are your thoughts about what's represented in the three figures *[Referring to Figure 3.1 a-c in the 'Neuron Assessment']*

ES: I think the diagrams show the basis of what the experiment is conducting. Figure 3.1c doesn't provide all the information it should. It's very minimal and basic. Figure 3.1a-b are much more specific and they show where everything is located in respect to the cells. So I think they depict whatever they are supposed to depict more efficiently.

AD: So what's going on in these figures according to you?

E: the…um...the axon transports in anterograde and retrograde directions.

AD: what tells you that something is getting transported?

E: in the third figure the arrows indicate movement and the labeling anterograde and retrograde also confirm the movement. Unfortunately in Figure 3.1a-b it doesn't exactly depict that. It just shows where the proteins are located in the cell.

AD: You mentioned the "axon transports". What indicates you see an axon?

E: Figure 3.1c is labeled axon.

AD: Where have you seen something like this before?

E: Not this exact process but in psychology I have seen similar types because you have to understand what neurons and axons work in the brain.

AD: So moving on the actual question, what are the scientists really trying to do here?

E: There are people with the disorder who are unable to perform transport that they need to and scientists believe that it has to do with the motor proteins-kinesin and dynein not working somehow and how that affects the movement of mitochondria

AD: What goal to these scientists have for this study then?

E: To determine if a problem with both, neither or one of the proteins *[kinesin and dynein]* is the source of the disorder and thus use that information to correct the process that is impaired in the disordered cells. So they want to fix that to make the neurons healthy in the person with the disorder to regain the movements that they need to carry out.

AD: So any idea how they would go about that?

E: the experiment?

AD: Sure. What would an experiment for this study involve?

E: Well you are going to need a control for an experiment [*Draws Figure 4C*]. The control will be the healthy neuron which has everything it needs to. Both neurons are going to contain the same organelles because that's required for the cell function. But experimental group will be the unhealthy neuron because we need to test that to find out about how the movement can be improved in the presence of kinesin and/or dynein. Control will just show the two proteins functioning normally.

AD: when you mention, "control and experimental group", what does that mean?

E: the control group is going to be everything you are in control of- so if you want a specific factor that you would like to maintain constant – that will be the control group. The experimental group is what you are going to add something to like the independent variable which you can decide how and how much of a variable is going to be added. Control is going to be set aside to see how things occur naturally and the experimental is you are going to decide how things occur.

AD: How would you use the materials provided in the study to actually perform your experiment?

E: the imaging software will help you record the movements that occur in the neurons. So you are going to use that for both control and experimental groups. The compound K and

D are inhibitors which will be injected in the experimental groups to see how they affect the neurons. You may go about doing the experiments separately like trying, just one compound and then the other or both together.

AD: How would you visually represent the different experimental scientists might try?

E: [Draws Figure 3.4D]

So you can try cells with just kinesin inhibitor, just dynein inhibitor and then kinesin and dynein inhibitor together. And then neither of them. With compound K injected, you are going to record what happens. For dynein you would inject compound D into the cell. If you want to see how the two proteins interact, you are going to inject both compound K and D.

AD: Why do you show 4 experimental groups and one control group in Figure 4D?

E: because they mentioned two proteins. The proteins could interacting or acting seepaerately. So one could have a hand in the process and the other couldn't or they could both be involved. They want to see how the proteins work in the cell and they also want to try it without them just to see how the process would be affected without any proteins.

AD: How will you decide the right samples for you each of your groups (columns in Figure 3.4D)?

E: for the expeirmental since you are injecting the compounds, you can use the same type of cell but you would inject different compounds. The control you want to use the healthy neurons to see how the process works in general or on its own.

AD: How will you present results of this study?

E: I think the most efficient would be graph. If they want to convey all the groups then they could use a bar graph showing the amount of movement or how many movements for a specific time period.

AD: Let's try and draw that graph maybe?

E: for the control there will be just one bar graph [draws Figure 3.4E].

Unfortunately since I don't know which protein has the effect I won't...be able…to…

AD: So let's imagine that nobody really knows and you are the one who gets to be the first one to find this out.

E: (*Referring to Figure 3.4E)* I am going to assume that both proteins have a hand in the moving of mitochondria. So the control shows how the process should occur normally. With the [presence of] proteins individually, they might have a little bit of effect on mitochondrial movement. But with both inhibitors together, that is going to have movements most close to the control. The x-axis is the proteins themselves. So the bars in

my graph show neurons with only functional kinesin (Graph b, bar 1), only functional dynein (Bar 2) and both functional proteins (Bar 3). And then the compounds are added to each kind of cell. I will then measure amount of mitochondrial movement with the imaging software although we don't have the healthy known amount of movements so you have assume that the control would provide the healthy amount of movements.

AD: So you think the control of a healthy neuron and healthy known amount of movements will be different in any manner?

E: I know it will be a little different in the unhealthy ones. So how the cells react is going to depends on how much you add, when and where you add it. Overall when you see movement for graph 2 (Figure 3.4E, right graph) closest to the control movement in graph 1 (Figure 3.4E, left graph), you would know that the experiment is successful.

AD: Tell me a little bit about you statement, "When you see movement for graph 2 closest to the control movement in graph 1, you would know that the experiment is successful."

E: The point of an experiment is to prove or disprove something to determine what you do is a success or a failure. Since we are saying that the control gives the healthy amount of mitochondrial movement needed, then with the experimental group you would want to find the group which is most closely related to the healthy. So whichever one is closest of the healthy, would determine what solution you would use to help the disorder.

AD: Summary in couple of lines

E: I want to determine which protein helps in solving the disorder. You would need to set up control and experimental groups- you would lay this out for the scientists. I would tell suggests the scientists use the bar graphs to determine compare your results because you want to pick the protein that's producing movement similar to the control.

Phase 3

AD: How would you rate these questions on a scale of 1-10? 10 being most conformable and 1 being I hope I don't have to ever do this again.

E: Since I have a basic understanding of how this experiments work, so I might be around 5-10 depending on which experiment. I will be honest because the third one is the more difficult one, I could more sufficiently explain the first 2 question set.

AD: tell me why was the third one relatively difficult?

E: Since I don't know v. much about the process in general and it would work, I feel my lack of knowledge in this topic didn't help me when I was answering this question. But the first two questions were much easier to understand because you only needed to understand how the experiment works to explain the context confidently. But in this question I was very skeptical of my own answers just because I don't have all the background information I need.

AD: So you think that the background info and figures provided did not make it easy for you to answer this question?

E: The background does sum up the basics. But I am kind of person where I want to understand it more sufficiently in order to explain it to somebody else or in order to come up with an experiment in my own sense. It is very difficult to come up with an experiment if you don't understand what you are supposed to find out eventually.

AD: Do you think any question about experiments is left out from what I asked you?

E: No I think all aspects are basically covered. I would expect going into science, you would understand the experiments generally because they teach you the scientific method. Usually we don't have to come up with our own experiments because all information in terms of how you need to set it up is provided. But you have to understand the basis like the control and experimental groups etc. to get there.

AD: If you were a diagram designer, would have drawn these pictures differently (Referring to Figure 3.1a-c the provided material)

E: Figure 3.1c has the basics but you kind of want to see how it happens. It would be great if that could be demonstrated. Figure 3.1 and 2 don't really show the process at all because it's just like here's everything in the neuron as its situated and here are the protein. So figures 1a-b really only help with understanding the cell set up. Figure 3.1c gives information of how the process occurs but may be you could have given a lot more.

AD: Do you think is question is clear enough for you?

E: If you ever want to go into a science career, that you are going to have to be able to make your own experiments and understand how to set them up and how to analyze results. These three questions really make you think about that-because in all our previous experiences, we were told how to do the experiment! We didn't exactly have to come up with our own and this really pushes you to gain that knowledge you need to set up an experiment yourself!

AD: How do you feel about participating in such exercises about experiments?

E: I feel they should try to do something like this into the courses because if you are always given the experiment and how to do it, you are never going to understand how you would make your experiment. That could hinder how you would approach an experiment in your own lab later as a researcher. These make you think about it and seek the knowledge you need to understand, the process and how you would set up a typical experiment, what you need, how would need the control and experimental. What do you record? I feel they should do something like this in the courses.

AD: Great! Thank you for participating!

E: Thanks !

3. Interview Transcript for Li Na

AD: Interviewer; Li Na (L): Student

Phase 1

AD: Hi! ST *[name hidden for confidentiality]*, I am Annwesa Dasgupta (AD). How are you doing today?

Li Na (L): Good!

AD: So, thank you for being here today. Alright, I would like to briefly explain some details about this activity. You just spent some time writing ideas about what you think about experiments. Now I would like to follow up your ideas by giving you an opportunity to share some thoughts verbally via a conversation based on a few questions. Are you ready to begin?

L: Yes!

AD: So, most of my questions will be related to the written survey you just completed, to help me understand your ideas. Some instructions to get started…please think freely about the questions I ask …there are no time limitations so you are free to take as much time as you wish to respond. There is no right or wrong answer to these questions. I am simply interested in your thinking about experiments. You are free to use provided materials to draw things if that helps you to express your thoughts. However, there might be certain instances where I request to visually present your ideas just so I am sure that I understand correctly. This interview will be recorded. You may choose to withdraw your participation at any time without penalty. If you have questions or need clarification at any time during this conversation, please let me know.

AD: So let's start with telling me what you according to you is a neuron?

L : Neuron?

AD: Ya!

ST: I know that neurons transfer signals and if you get signal from outside of the body like someone touches you or you hear something, the neuron can transmit that information to your brain.

AD: How would share that in a drawing? Also please label your diagram.

L**:** *Draws 3.5A*

AD: This is a nice drawing (referring to Figure 5A). This is your drawing about neurons. Now if I ask you what you think about "organelles moving inside of neurons", what would you say?

L: before this survey I just knew about how neurons communicate with each other and how the gradual change in ions across a membrane help in transmitting signals along axons (as drawn in Figure 3.5A). I only know about this aspect but I don't know anything about mitochondria transportation.

AD: Ok you drew this figure of a neuron. Can you picture mitochondria in the neuron anywhere?

L: maybe just along the axon (Draws and labels *mitochondria* in Figure 3.5A).

AD: how did scientists discover the ideas you share in your nicely drawn Figure 5A?

Li Na: They might have labeled the important organelles.

AD: So you mention "labeling". Tell me a little more about that?

L: maybe somehow they would amplify the process and label some important organelles. They could explain it in words instead of drawing it because they might not know how the process looks.

AD: Would they have made any measurements?

L: So if we consider that scientists know the structure of organelles but they are not sure how they move, they could measure the direction and displacement or electrical potential.

AD: Under what conditions would they have made these measurements?

L: Might have labeled the important organelles. Also the presence of different amounts of ATP present might affect the directions in which the organelles move.

AD: Any idea how they would they have actually carried out what you suggest?

L: They would have to use a computer program because they organelles are really small. I don't think you can they can be recorded using naked eye.

AD: How would summarize your ideas in a couple of sentences to explain your ideas on what scientists would do to measure movements along a neuron?

L: I don't know how to explain it. Let me try. I would first set up a hypothesis.

AD: What would that hypothesis be then?

L: The scientists want to measure which organelle will cause movement in different directions. After the hypothesis, they will set up an experiment.

AD: How would they go about that?

L: they know how the organelles move but they don't know *[pause]*…they know the structures and the movement are based on myosin. They consider other variables that would cause a difference in the direction of movement.

AD: When you mention variables, what do you mean?

L: You need to change certain things and not just observe them. After that, you get different responses from variables in an experiment.

Phase 2

Probe for surface-level reasoning

AD: Now! Here is a sheet with couple of figures that are the same figures you saw in the written survey you just completed. Along with these figures, here is another sheet with some background information. I would request you to take some time to go through these sheets. Let me know when you are ready. *[Showed the figures to the participant...gave them some time to think about what they are seeing…and followed up with these questions below...]*

L: *[after few minutes]*… I am ready. I just went through these sometime back so I am familiar with these.

AD: What are your thoughts about what's represented in this diagram?

L: In Figure 3.1a, I know that the mitochondria are along the axon of a neuron and I can compare Figure 3.1a and 1c. I find Figure 3.1c an easy one. I also see a cell nucleus and cell body. Figure 3.1c is more easily understandable but Figure 3.1 gives a more accurate structure. Figure 3.1b is an amplification of Figure 3.1a.

AD: So what's going on in these figures?

L: I know the kinesin and dynein can cause movement in different directions of mitochondria because I see arrows in Figure 3.1c which tells me about a difference in directions. Figure 3.1b is really different. I see microtubules around the mitochondria but in Figure 3.1a I don't really see microtubules. I also notice that a difference in calcium ions cause a difference in direction. So ions interaction causes a difference in direction.

AD: so you mentioned this is 'neuron'. What do you think so?

L: from the structure in Figure 3.1a which is really representative of a neuron.

AD: what features of a neuron do you see here?

L: different terminals like cell terminal and there is a cell body.

AD: Where have you seen a neuron before?

L: just in the textbook from my course before.

AD: you mentioned about "movement in different directions"? What tells you that you see movement?

L: I see myosin and ATP which I guessed indicates an energy change and movement.

AD: What are scientists trying to do in this study?

L: they are trying to the find the cause of a disorder.

AD: Tell me a little more about that.

L: The disorder may bring pain to the patients so they are trying to find a way to cure them. The transportation in the anterograde and retrograde directions are both activated because kinesin and dynein are both active. So the mitochondria cannot move in either direction because the kinesin and dynein cancel each other-and so this maybe the disorder.

AD: How would they use the materials provided to study the cause of the disorder as you just mentioned?

L: they might try four combinations (outlines in Figure 3.5B) as treatments for the mitochondria.

AD: When you mention treatments, what do you mean?

L: Treatment… *[pause]*..Before the treatments the subjects should have the same conditions and then you try different things on them and see the response.

AD: Tell me more about what you mean when you say, "Before the treatments the subjects should have the same conditions"?

L: if they don't have the same conditions, they may react differently and that may lead us to think about false causation.

AD: So what kind of conditions would you keep the same in this study you are proposing?

L: I will keep the same organelles under observation, use the same species of organisms for the neurons and use cells from the same one animal. And also make sure that they are in the same environment.

AD: So you mention 4 combinations? Why so?

L: for an experiment, they need to find a cause and for that they need to set up control groups and experimental groups. We are given two compounds, a kinesin and a dynein inhibitor and by inhibiting we can look for effect on neuron function.

AD: What does a control group mean to you?

L: The baseline. I cannot remember the exact concept. But you need a control group to come the experimental groups to it.

AD: how will you decide the right sample for the treatment and control group?

L: the sample/subject is the mitochondria in the neuron and kinesin/dynein is the variable because they will be either inhibited or activated. If in the control group, displacement of mitochondria in either direction is zero.

AD: What kind of results do the scientists expect to get from the combinations you suggest?

L: the kinesin moves mitochondria in the anterograde direction while dynein moves it in the retrograde direction. Both if activated together will result in the disorder. Then I will measure the direction and displacement and draw a graph like this (*Draws Figure 3.5C*).The y-axis will show the displacement and x-axis shows "+" for anterograde movement and "-" for retrograde movement. Group 1 is the control group. Group is activated kinesin and inhibited dynein so we see only anterograde movement. Group 3 is both activated. Group 4 dynein active and kinesin inhibited so the movement is in retrograde direction.

AD: In what format will the results be recorded?

L: I think the results should be recorded in form of numbers. Maybe displacement can be measured in terms of length in micrometers.

AD: If you had to go back and summarize the overall experiment you designed from beginning to end in a couple of sentences, what would you say? If it helps you can also visually represent your experimental proposal.

L: First I would have a hypothesis. Then do the experiments. Then show the results. When kinesin is activated and dynein is inhibited, we see movement in the anterograde direction. When dynein is working and kinesin is inhibited we see movement in the retrograde direction. When both are activated, the functions of the two proteins are replicated and thus, the mitochondria cannot move in either direction so the movement is impaired.

AD: You mentioned replication. What does replication mean?

L: when a large number of samples are used to avoid the chance variable.

AD: What is a 'chance variable'?

L: I have just learned this few weeks ago. Having small groups might lead us with results that are not persuasive. If you get a larger number of samples, you can see the outliers of the data clearly and then just pick the values that lie centrally.

AD: How would you increase the validity of your experiment?

L: by using randomization. When you choose the samples, you assign them randomly.

AD: Describe that a little more.

L: cells even when taken from one animal might have differences. So when you extract them you need to pick them randomly and then also randomly assign them to the experimental groups. People might do that to decrease the confounding variables-so if one group has a special tendency for a certain kind of trait; they will react and lead us to wrong causation. So randomization is very important.

Phase 3

AD: How would you rate these questions on a scale of 1-10? 10 being most conformable and 1 being I hope I don't have to ever do this again.

L: I would say 9.

AD: Tell me why?

L: I think I can come up with a lot of ideas so I am comfortable with activities like this.

AD: Is there anything in particular about this question that you don't quite understand or find confusing?

L: yes. In Figure 3.1b, I see calcium ions but I am confused about the roles of that.

AD: Do you think any question about experiments is left out from what I asked you?

L: yes! How are the kinesin activated or inhibited? What causes their activation or inhibition? Most of the people usually don't carry this disorder so one functions then…but I think both are present in neurons structurally. But how can they be selectively activated or inhibited? I am not sure how the compounds cancel each other.

AD: if you were a diagram designer, would have drawn these pictures differently (Referring to Figure 3.1a-c in the question material)

L: I am confused about how the mitochondria are outside the microtubule. Also I will label ions for dynein.

AD: Do you think is question is clear enough for you?

L: Maybe. I don't know the answer to this experiment so whether the question is good depends on the answer.

AD: How do you feel about participating in such activities about experiments?

L: Maybe it's good for future. I find it interesting!

4. Interview transcript for Daniel

Interviewer: AD Student: Daniel

Phase 1

AD: Hi! DW [name hidden for confidentiality], I am Annwesa Dasgupta (AD). How are you doing today?

Daniel (D): Good!

AD: So, thank you for being here today. Alright, I would like to briefly explain some details about this activity. You just spent some time writing ideas about what you think about experiments. Now I would like to follow up your ideas by giving you an opportunity to share some thoughts verbally via a conversation based on a few questions. Are you ready to begin?

D: Yes!

AD: So, most of my questions will be related to the written survey you just completed, to help me understand your ideas. Some instructions to get started…please think freely about the questions I ask …there are no time limitations so you are free to take as much time as you wish to respond. There is no right or wrong answer to these questions. I am simply interested in your thinking about experiments. You are free to use provided materials to draw things if that helps you to express your thoughts. However, there might be certain instances where I request to visually present your ideas just so I am sure that I understand correctly. This interview will be recorded. You may choose to withdraw your participation at any time without penalty. If you have questions or need clarification at any time during this conversation, please let me know.

AD: let's talk a little bit about neurons. What's the first thought that cross your mind when I mention "neuron"?

D: like nerves.

AD: tell me a bit more about that…

D: Just like signals throughout your body-signals to move or other processes.

AD: If you had to draw a nerve, what you would draw?

D: something like...I guess [Draws Figure 3.6A] a tree. So you start with a thicker nerve and then it branches off, into smaller and smaller pieces, until it gets to the end…

AD: Would you label any parts?

D: I don't really have anything to label.

AD: Ok! Building on this figure of a neuron, when I say organelles moving along neurons, what would you say?

D: uhh…I don't know I just think of electrical signals. Other than that I don't have any information.

AD: you mentioned, "Electrical signals". How would you depict that in this figure?

D: Umm I don't know. I would assume it would move against the wall of the neuron [Figure 3.6A].

AD: How did the scientists' find out about the things like electrical signals along neurons etc.?

D: I would assume some sort of experiment involving people with impaired nerves or something along that nature. Then comparing that to like a control group with others that have normal/regular nervous system.

AD: How would you schematically depict what you just mentioned?

D: [*Draws* Figure 3.6B] So you have a control carrying people whose nervous system isn't impaired. Then you would have to compare signals among people in the control groups with people in the experimental group that have an impaired nervous system.

AD: When comparing signals [Figure 3.6B], would the scientists' be measuring something?

D: I am sure they would be measuring something because they probably should be something that could be measured. You could measure the strength of the electrical signals or the path the signal takes and see differences in the way a normal person's body would send signals out vs. somebody with an impaired nervous system. And how the body responds…

AD: Would there be any numbers involved?

D: If that's possible. That's probably the best way to do it. But I am not sure…

AD: Under what conditions would they be making these measurements?

D: they would probably have two similar types of people with as little different between them except for the nervous system.

AD: Why do you suggest that?

D: people that are of different height would either send weaker/stronger signals because of the distance they would have to travel. Age might affect it. So the two types of people should be very similar except their nervous system.

AD: You mentioned great ideas to suggest what scientists would have probably done to find out about electrical signaling along neurons. How would you summarize in 3-4 short sentences?

D: scientists would try to measure the electrical signals in the two different groups: 1) control group with normal nervous system. 2) Another group that would have the nervous system impaired in some way and they would compare the signals/path/strength or something like that in the two groups. They would try to keep those as similar as possible so it's just the nervous system that's different between the two so the results aren't affected.

AD: Results aren't affected means what?

D: I mean if there is a difference between heights of subjects in two different groups, you wouldn't be able to necessarily decide if it was the height that gave rise to the difference in strength of the electrical signals rather than the nervous system.

Phase 2

Probe for surface-level reasoning

AD: Now! Here is a sheet with couple of figures that are the same figures you saw in the written survey you just completed. Along with these figures, here is another sheet with some background information. I would request you to take some time to go through these sheets. Let me know when you are ready. *[Showed the figures to the participant...gave them some time to think about what they are seeing…and followed up with these questions below...]*

D: [After couple of minutes] I think I am ready now…

AD: Great! So first, what are your thoughts about what's represented in the three figures [Referring to Figure 3.1a-c in the 'Neuron Assessment']

D: the mitochondria moves through the axon in Figure 3.1a which sends some sort of signal and then its moved using the two proteins [kinesin and dynein].

AD: You mentioned that mitochondria moves? What in the figures gives you an indication of movement?

D: I'd say the arrows on Figure 3.1c show that one protein goes one way and the other goes the other way. They move along an axon of a neuron.

AD: What tells you that you see a neuron?

D: I don't know. I think just because it said in the part of the question. But it also kind of looks like what I drew earlier so I think I am familiar with a similar structure of the neuron.

AD: Cool! Have you seen figures like this before?

D: I don't know about this stuff specifically but I know like biology classes in high school they have shown more basic figures of what nerves looks like without the more detailed explanation.

AD: What are scientists' trying to do in this study?

D: they think the two proteins help in the movement and some disorder is caused they believe by the proteins not doing what they are supposed to. This causes the mitochondria to not move how it should. They are trying to determine first of all, if actually these are the proteins that help movement and then want to determine if those are what's wrong with people who have the disorder.

AD: So do they have a goal in this study?

D: to find out which of the two proteins causes the disorder so that they could try to fix it?

AD: What ideas do they have in terms of that goal?

D: They have two different compounds to inhibit the two different proteins and observe which inhibited protein affects mitochondrial movement in a manner similar to the movement in people with disorder. They also have software to measure the movement with those who has the protein inhibited or when they are not. They can use the imaging software and determine the movement with the inhibited proteins and see if it's similar to the movement in those who have the disorder.

AD: Let's imagine you are the lead scientist of a group that is supposed to conduct the experiment you just described. What specific directions would you give your team to carry out this experiment using the materials provided? Also try maybe depicting it in some form of a visual like a schematic or flowchart.

D: Ok it might be easier for me to think about it and draw something first.

AD: Sure go ahead; take your time to draw ideas.

D: [Draws Figure 3.6C]

AD: Can you please walk me through your diagram [Figure 3.6C]?

D: Ok so I started out with measuring movement of mitochondria in nerves of a normal person. Then I split a group of normal people's cell cultures into four different groups, control groups, one with compound K, one compound D and one with both. I am assuming these people were similar to each other as much as possible, in like their health conditions, such that we know that the observed effect is due to the application of compound K or D. Then you could measure the movement in each of those groups. Then compare the movement with multiple patients who have the disorder with the 4 groups of patient. This will allow us to infer that those were the protein that caused the disorder.

AD: What does a control mean to you in an experiment?

D: I guess a group that would not be receiving any treatment but other than that it would still be subjected to the same conditions as those who are given the treatment (compounds in the case of this study).

AD: Tell me why do you have 4 groups here (Figure 3.6C)?

D: The control group allows them to measure changes in the movement while the experiment was going on. Just the K and D because those are two things whose effect will be measured. I figured I would test both in case the patient had both that weren't working correctly. Then you would also have to have the group of patients [with the disorder] to be able to test to see if the difference in their movement was the same. So they would know what they found in their experiment is actually what is wrong with the patient.

AD: how will you decide what kind of patients participate in your control vs. other groups with compounds applied?

D: I would randomly assign them into groups. Like I would number each patient and use a random no. generator…so for example, if this was out of a 100 people, the first 25 are placed in the control, the second 25 in the next group and so on….

AD: What is the relevance of "randomly assigning" as you mention?

D: if you just grouped them in a non-random manner it wouldn't be evenly spread out between all the different variables. If you did it by height, you would bias the results and find differences across groups due to the height differences rather than a result of compound application.

AD: What kind of results do scientists expect to get? What would those mean?

D: Like before I will try drawing it out [draws Figure 3.6D].

So…I just made up different numbers they might have gotten as results although I am not sure of the units on it. Then just take the patient with disorder and if it matched around the same range as movement in the compound D, they would know that a problem with dynein is the cause of the impaired mitochondrial movement.

If it was a different number, they would know a problem with those compounds have no role to play in causing the disorder.

AD: How would scientists visually represent these results? How would they communicate their results to another group of scientists?

D: They would probably present a report with graphs.

AD: How would they draw that graph?

D: [Further adds to Figure 3.6D]

AD: How would you explain this graph to me?

D: I listed the different treatments on the x-axis. Along the y-axis is the movement compared to the control group. I would just graph the difference in movement from one to the other. Then you would compare to see how similar are the differences with the treated cells to the actual cells from patients with the disorder.

So the first bar shows that with treatment with compound K, the mitochondria moved 4 units less than the control groups it over a specific period of time. And so because treatment with compound D moved 6 UNITS less than the control group, dynein inhibition more strongly affects overall mitochondrial movement. Alternatively, you could also just graph a bar for the control group and compare them.

The scientists could then develop something to make the protein work or fix the existing problem somehow.

AD: How would summarize your experiment in 3-4 lines?

D: 1. Measure movement of mitochondria in neurons for a group of randomly picked normal persons who are as similar to each other as possible in terms of general health conditions.

2. Split cells of normal persons into 5 different groups. Each group carries a different treatment as outlined in Figure3 [normal person; control with no treatment, one with compound K and another one with compound D; one gets both]

3. Compare your movement with the treatment groups to the movement in neurons of a patient with disorder to see if there are any similarities in trends of the movement. If they did have the same movement, you could argue the source of the disorder as per your treatment.

Phase 3

AD: How would you rate these questions on a scale of 1-10? 10 being most conformable and 1 being I hope I don't have to ever do this again.

D: I was pretty comfortable with the way the questions were framed so I would say 9.Only thing I wasn't so sure about was not knowing more background information when designing experiments or answering questions. Just not being sure what exactly might be affected in the real patients.

AD: is there anything in particular about this question that you don't quite understand or find confusing?

D: The only thing I found confusing was Figure 3.1b which was little busy.

AD: Did the diagram and background information, help you, in thinking about your ideas?

D: The diagrams definitely helped me think about the process more clearly since I did not know about this process too much before this study. I think it helped me see how things like the mitochondria, kinesin, and dynein are placed within a neuron.

AD: Do you think any question about experiments is left out from what I asked you?

D: I don't think so….

AD: If you were a diagram designer, would have drawn these pictures differently (Referring to Figure 3.1a-c)

DW: I don't know about changing them but most textbooks have a couple of sentences explain each figure. Including something like that might be helpful to better understand the process of what's going on.

AD: Overall do you think this is a clear question?

D: yea it was pretty good. I like it. After reading all the provided material it was easy to understand what information they already had and what they are not looking for.

AD: What is your take answering such question in general? How do you like the process of figuring out about experiments in a format that you just participated in?

D: I liked it! It was quite okay. So far in biology we haven't really had to come up with our own experiments. It's more of we were asked to read what other people had done and their experiments and how they dealt with different things. It's nice and probably important to be able to think through what you would do as a scientist. This pushes me to decide about things I haven't thought of before.

Appendix L: Written Assessment Responses

Expert

The questions on this page do not require much scientific knowledge. Most scientists can answer these questions. Please write down your own ideas in response to these questions. There is no time or word limit so feel free to share your thinking as much as you can. Please ask for more writing paper if you need.

Question 1: The "Shrimp" Question

Acknowledgements: The College Board (2006) AP® Statistics Free-Response Question 5
[Online http://apcentral.collegeboard.com/apc/members/exam/exam_questions/8357.html]
(Used with permission to Nancy Pelaez, npelaez@purdue.edu)
Background Information

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

a) Consider the treatments that the biologist plans to use in this experiment. Create a visual representation and explain the treatments that the biologist plans to use.

b) Using the treatments listed in part (a); describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks. What visual representation will allow the biologist to compare the shrimps' growth after 3 weeks?

c) Consider one statistical advantage to having only tiger shrimps in the experiment. To explain why this is an advantage, create a visual representation that illustrates one statistical advantage to having only tiger shrimps in the experiment.

There will be species uniformity.

Some Sources of variability
Tank temp
Measurement error
Food
Catch date
Species ← Remove this as a source
Animal age

d) Consider one statistical disadvantage to having only tiger shrimps in the experiment. To explain why this is a disadvantage, create a visual representation that illustrates one statistical disadvantage to having only tiger shrimps in the experiment.

This limits the generality of the conclusions of the effects of A.

Animals
Non-shrimp
Shrimp
Tiger shrimp
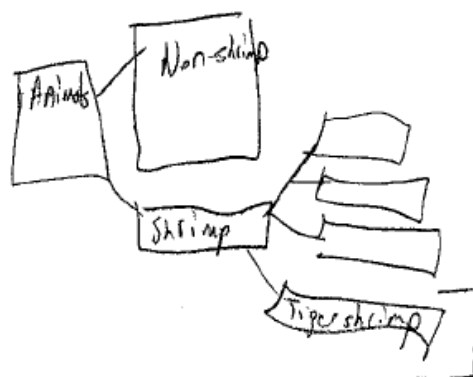
One can only draw conclusions about this species under these conditions

Question 2: The "Drug" Question

√ = I would not typically use visual representations here.

Background Information

The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing.

Directions

As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

a) Using complete sentences state the hypothesis to be tested. Create a visual representation to explain your hypothesis.

Alamain will lower blood pressure in humans with high blood pressure.
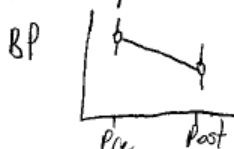
b) Since there are several contributing factors that can affect blood pressure levels, list **five** factors that will be constant between the experimental and control groups. Create a visual representation that illustrates these constant factors between experimental and control group. *?*

Factors: Age, Gender distribution, socioeconomic status, initial blood pressure, initial weight
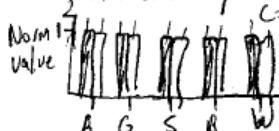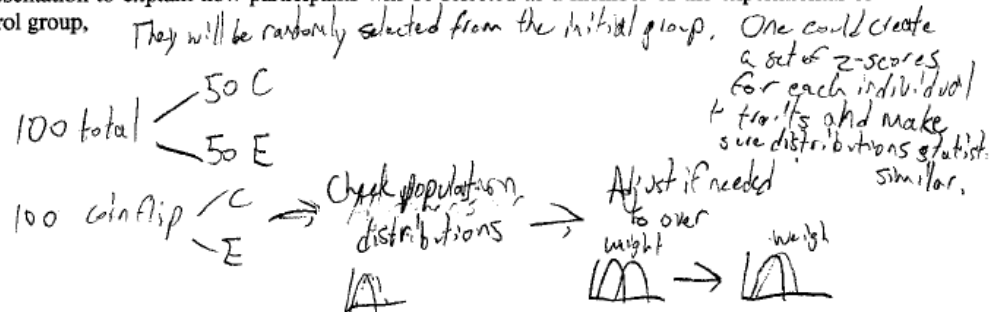

c) Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study. What visual representation will the scientist use to explain your answer?

Which scientist?

The goal is for the control and experimental populations to be statistically equivalent in as many relevant factors as you can think of and control for. In this way, one hopes to isolate the drug effects without confounding factors (e.g. only works for younger patients or only for men).

d) Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group. Create a visual representation to explain how participants will be selected as a member of the experimental or control group.

They will be randomly selected from the initial group. One could create a set of z-scores for each individual + traits and make sure distributions statist. similar.

100 total — 50 C / 50 E

100 coin flip — C / E → Check population distributions → Adjust if needed to over weigh / we igh

e) Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken. What visual representation will allow the scientist to make measurements to judge the efficiency of Alamain?

I have no idea what is standard, so I'd consult medical or previously published standards. Also, no time frame is given. Is the drug acute or chronic?
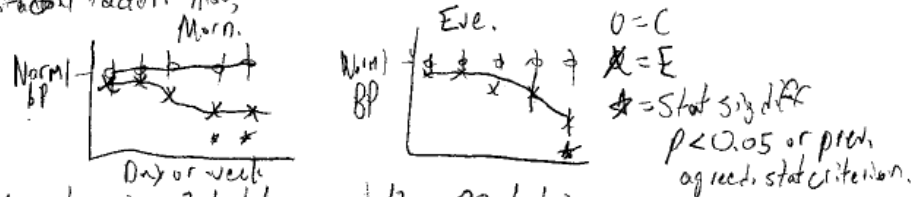
BP measurements

Let's say, since BP measurements are simple, daily morning (just after wakeup) + evening (before bed) for 8 weeks
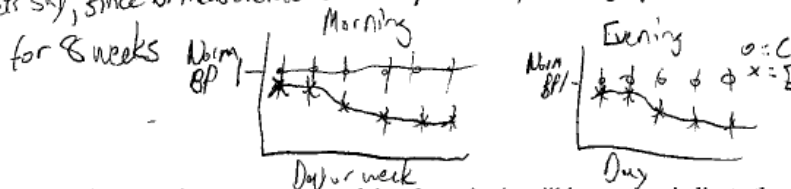
Morning — Norm BP — Day or week

Evening — Norm BP — Day — o = C, x = E

f) Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans. What visual representation will allow the scientist to indicate the success or failure of the drug Alamain?

Again, need to know chronic or acute for time frame.

I would probably do a repeated measures ANOVA to test for efficacy, looking for treatment group as a significant factor. Also,
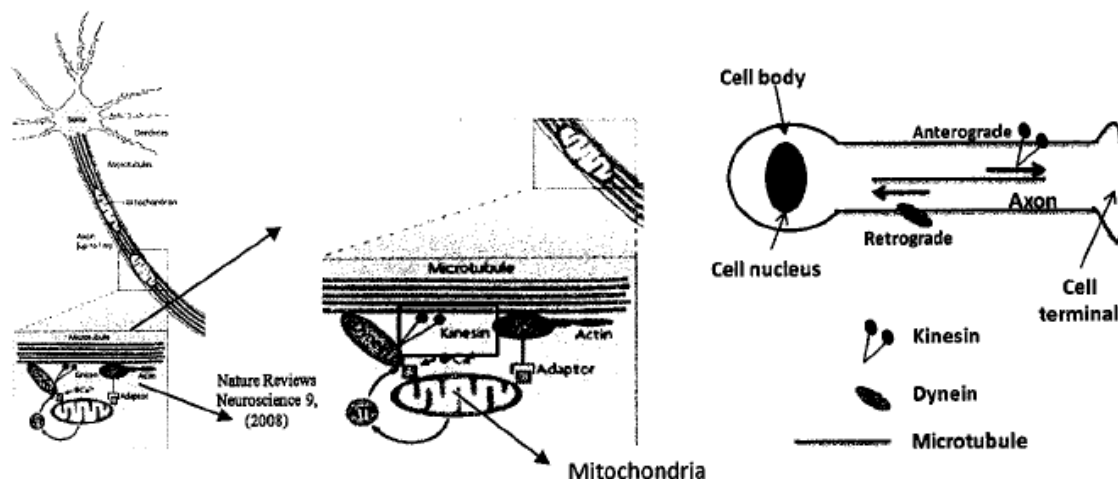
Morn. — Normal BP — Day or week

Eve. — Norm BP

o = C, x = E, ☆ = Stat sig diff, p < 0.05 or prev. agreed stat criterion.

Could also do an X week post-test to determine whether effects last

Question 3: The "Neuron" Question

Acknowledgements: Original question by Annwesa Dasgupta

Figures

Nature Reviews Neuroscience 9, (2008)

Background

Mitochondria are one of the several organelles that get transported across the axon of a nerve *(Refer figure above)*. They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions

Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder caused by impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons.

The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;

Compound D: inhibits dynein;

An Image software: measures mitochondrial movement in neurons.

- How do you think a 'hypothesis' relates to an experiment? Create a visual representation to illustrate your answer. *An experiment tests a prediction that arises from a hypothesis.*
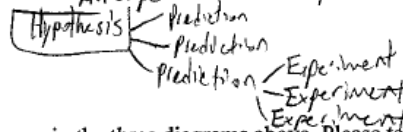
[Hypothesis] — Prediction
— Prediction
— Prediction / Experiment
— Experiment
— Experiment

a) Describe what you see in the three diagrams above. Please tell us in detail what you think about it. *Mitochondrial transport in axons via the molecular motors kinesin + dynein is depicted. Details of the transport process are along the microtubule are shown. +Myosin It suggests transport is a Ca²⁺ ATP dep. process. Dynein transports retrogradely along microtubules + kinesin anterogradely.*

b) What could be a potential hypothesis for your experiment? Create a visual representation to illustrate your hypothesis. *At this point, there is at least a vague suspected directionality for the experiments, which is mostly set up as observational (Let's see what happens). However to make a more specific hypothesis. Given the prevalence of mitochondria in axon terminals, kinesin transport of mito. will predominate over dynein.*
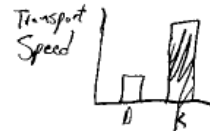
Transport Speed [bar graph: D low, K high]

c) Which factors will the researchers vary and which will they keep the same in this study? Why? Use a visual representation to explain the factors they will vary and keep the same.
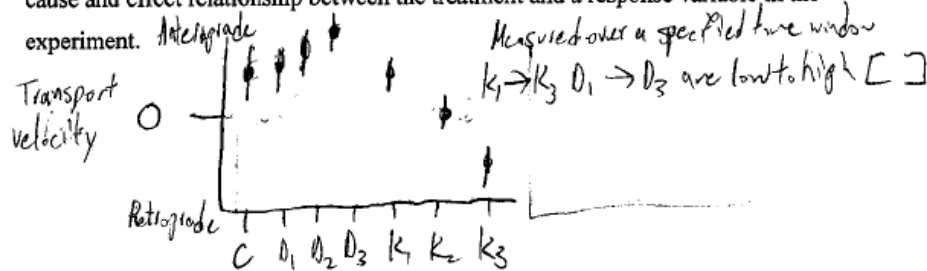
[ATP] Same
Cell type
Growth conditions
Culture age
Temp
Observation times
Analysis

Vary
[K]
[D]
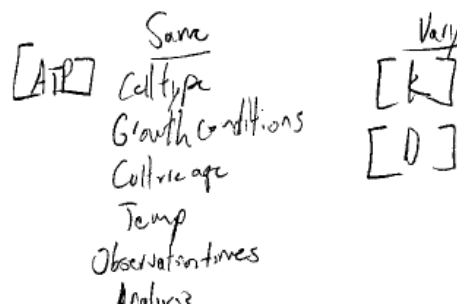
d) Consider how researchers will assign subjects to groups for this experimental study. To explain how they will assign subjects, create a visual representation to support your answer. Depends how easy it is to apply k or D focally. Generally, one would apply k or D to a culture dish + observe effects. Subjects are a set number of neurons in dish (10?)

$3[k]$
$3[D]$
1 baseline

$(k_1)(k_2)(k_3)(D_1)(D_2)(D_3)(base) \times 5$ replicates

e) Consider a cause-and-effect relationship between the treatment and a response variable in this experiment. To justify this relationship, create a visual representation to explain a cause and effect relationship between the treatment and a response variable in the experiment.

Anterograde

Transport velocity

Retrograde

C D_1 D_2 D_3 k_1 k_2 k_3

Measured over a specified time window $k_1 \rightarrow k_3$ $D_1 \rightarrow D_3$ are low to high [ ]

f) How would the researchers present the results of this experiment?

Many possible ways. Graphically, as above. In a movie form, showing differential rates of transport for different treatments, etc.

g) What results do the researchers expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of this experiment. What visual representation will allow them to present results?

Success or failure? I suppose failure is lack of effect of k or D?

The researchers expect to get a series of images of mitochondrial positions for each axon tested. From this, they can measure mitochondrial velocity or position as a function of time. I would probably use the transport velocity above and possibly a time series of positions.

h) Consider improving the validity of the experiment. What visual presentation will you use
   to show how will the validity be improved? *Not sure what this means. Does this mean:
Scope of inference? If so, the perform same experiment on a different neuron type
perhaps?*

i) What do you think this diagram is not showing? Explain your answer.
   It is not showing the high density of mitochondria in terminals.

j) Is there anything about this question that you don't understand or find confusing?
   Explain. *There is not really an associated hypothesis at this point. It shows
   dynein and kinesin simultaneously bound to microtubule*

k) Consider yourself a diagram designer. If you could change the diagrams, what would you
   change or how would you improve them? *I might show the mitochondria - motor - microtubule interactions separately first
   before showing them all together.*

Juan Written Assessment Response

The questions on this page do not require much scientific knowledge. Most scientists can answer these questions. Please write down your own ideas in response to these questions. There is no time or word limit so feel free to share your thinking as much as you can. Please ask for more writing paper if you need.

Question 1: The "Shrimp" Question

Acknowledgements: The College Board (2006) AP® Statistics Free-Response Question 5
[Online http://apcentral.collegeboard.com/apc/members/exam/exam_questions/8357.html]
(Used with permission to Nancy Pelaez, npelaez@purdue.edu)
Background Information

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

a) Consider the treatments that the biologist plans to use in this experiment. Create a visual representation and explain the treatments that the biologist plans to use.

The biologist will first begin by creating a control, a tank with 10 shrimps and no growth-enhancing nutrients and normal salinity levels. Then one will split the other 11 tanks, each containing a certain nutrient or salinity level, or a mixture of the two; for example, one tank may include normal salinity level and nutrient A, whereas another tank may contain nutrient A and high salinity levels.

control	only A	only B	only C	no nutrients only low salinity	no nutrients only high salinity	A + low salinity	A + high salinity	B + low salinity	B + high salinity	C + low salinity	C + high salinity

b) Using the treatments listed in part (a); describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks. What visual representation will allow the biologist to compare the shrimps' growth after 3 weeks?

The biologist, presumably measured the average size of the 10 shrimps in each tank so as to get the average initial length, or the biologists could mark each shrimp in a different way and measure their length at the start of the experiment. After 3 weeks, (s)he could measure each shrimp size and compare it to the starting size. Initial lengths may vary. For the mark, he could tie a different colored ribbon to each shrimp.
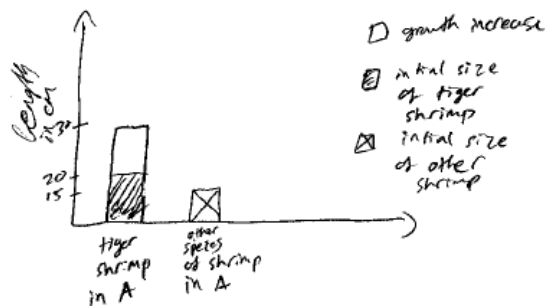
c) Consider one statistical advantage to having only tiger shrimps in the experiment. To explain why this is an advantage, create a visual representation that illustrates one statistical advantage to having only tiger shrimps in the experiment.

Tiger shrimps operate in the same way, they make use of the same nutrients and react the same way to all products. Having only tiger shrimps, makes it a controled factor, so it any shrimps did grow, the biologist would know that it was not because of the time factor, but because of the variable factors in the tank. Graph below is an example of what might happen in 3 week!
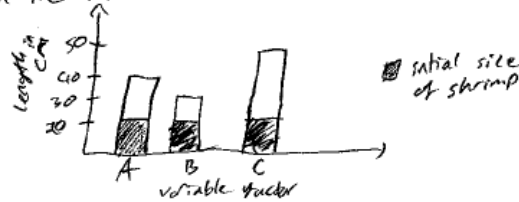
d) Consider one statistical disadvantage to having only tiger shrimps in the experiment. To explain why this is a disadvantage, create a visual representation that illustrates one statistical disadvantage to having only tiger shrimps in the experiment.

Tiger shrimps may contain a hormone that other shrimps lacks, that help digest these certain growth-enhancing nutrients, and using only tiger shrimps, you are finding a solution to only one problem (tiger shrimp species).

Question 2: The "Drug" Question

Background Information

The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing.

Directions

As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

a) Using complete sentences state the hypothesis to be tested. Create a visual representation to explain your hypothesis.

ALAMAIN will lower the blood pressure in people whose blood pressure is too high.

b) Since there are several contributing factors that can affect blood pressure levels, list **five** factors that will be constant between the experimental and control groups. Create a visual representation that illustrates these constant factors between experimental and control group.

① Food consumption between patients would be the same, same diet.
② Activities like excercise or work would be the same between the patients.
③ The amount of the drug will be given evenly to each patient.
④ The patients are to be treated the same way, no individual attention.
⑤ All patients must have high blood pressure.
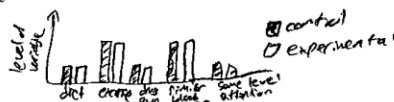
c) Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study. What visual representation will the scientist use to explain your answer?

Scientist should keep a log on what they eat as well as how many hours they work, they should control excercise activity and administer the same amount of drug to each patient, All these criteria eliminate the other variable factors leaving only the hypothesis as the question that will be answered. If any side effects occur in only one patient, we could further research into

1

d) Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group. Create a visual representation to explain how participants will be selected as a member of the experimental or control group,

Ideally, we would like the members of the controlled group to be similar in physical and medical shape to members of the experimental group. We would tell each group that they would be taking the drug, whereas the control would be taking a decoy.

e) Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken. What visual representation will allow the scientist to make measurements to judge the efficiency of Alamain?

A tablet will be given to both groups every day, without telling which group is the control and which is the experimental and their blood pressures, would be measured daily/weekly in order to observe the difference in each group.
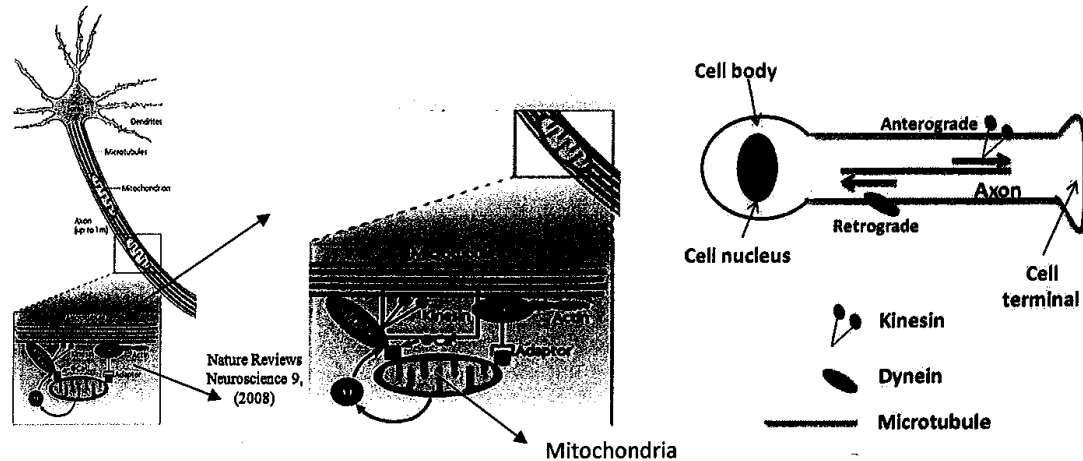

f) Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans. What visual representation will allow the scientist to indicate the success or failure of the drug Alamain?

Depending on the side effects uncovered, one might find that the blood pressure did indeed go down, but if the side effects cause other major problems then the drug is most probably a failure.

Question 3: The "Neuron" Question

Figures

Mitochondria

Background

Mitochondria are one of the several organelles that get transported across the axon of a nerve *(Refer figure above)*. They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions

Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder caused by impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons.

The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;

Compound D: inhibits dynein;

An Image software: measures mitochondrial movement in neurons.

- How do you think a 'hypothesis' relates to an experiment? Create a visual representation to illustrate your answer. A hypothesis gives an experiment a purpose, becau an experiment is done to receive an answer to the hypothesis.

a) Describe what you see in the three diagrams above. Please tell us in detail what you think about it. In the first figure, I see an axon and an enlarged image of one of the mitochondria in it. In the second figure I see an enlarged image of one mitochondrion. In the third diagram. I see a summary of the process being talked about. These three diagrams summarize what the passage says and helps the reader further understand the concepts being mentioned.

b) What could be a potential hypothesis for your experiment? Create a visual representation to illustrate your hypothesis. Does kinesin or Dynein play a role in the cause of the disease.

c) Which factors will the researchers vary and which will they keep the same in this study? Why? Use a visual representation to explain the factors they will vary and keep the same. They will have a control that will have both compound K and D, one with only compound K and another with only compound D. That way we can figure out what the causing reagent is.
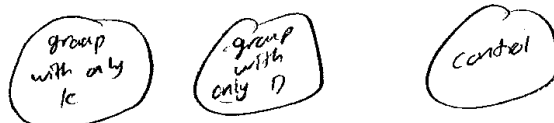
group with only K

group with only D

control

d) Consider how researchers will assign subjects to groups for this experimental study. To explain how they will assign subjects, create a visual representation to support your answer. *They will randomly an equal amount of people with the same disorder and place them into each group.*

e) Consider a cause-and-effect relationship between the treatment and a response variable in this experiment. To justify this relationship, create a visual representation to explain a cause and effect relationship between the treatment and a response variable in the experiment. *The lack/substitution of compound k with a genetically engineered compound may lead to the patients getting better or vice versa. Same with compound D.*

f) How would the researchers present the results of this experiment? *They would explain which compound (if any) proved to be the initiating factor and they would justify with the experiment and visuals.*

g) What results do the researchers expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of this experiment. What visual representation will allow them to present results? *They would expect that one of the compounds was indeed the initiating factor to the disorder. The experiment would be a success if the patients condition improved, a failure if the patients conditioned stayed the same or deteriorated. The image software.*

h) Consider improving the validity of the experiment. What visual presentation will you use to show how will the validity be improved?

The image software that measures mitochondrial movement in neurons will be used.

i) What do you think this diagram is not showing? Explain your answer.

What kinesin's and dynein's functions are.

j) Is there anything about this question that you don't understand or find confusing? Explain. If we remove kinesin or dynein, will the patient suffer any side effects (is it possible to remove one, without replacing it).

k) Consider yourself a diagram designer. If you could change the diagrams, what would you change or how would you improve them?

I would focus more on the kinesin and dynein.

Eve Written Assessment Response

The questions on this page do not require much scientific knowledge. Most scientists can answer these questions. Please write down your own ideas in response to these questions. There is no time or word limit so feel free to share your thinking as much as you can. Please ask for more writing paper if you need.
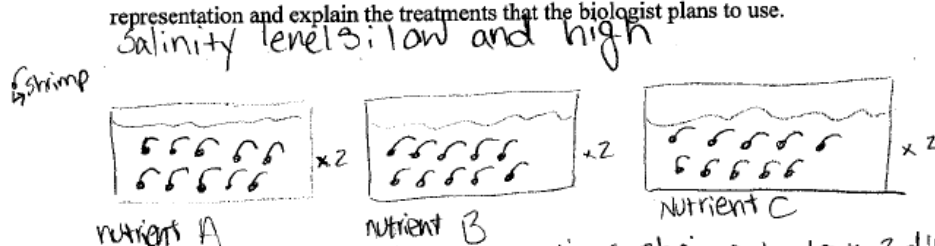
Question 1: The "Shrimp" Question

Acknowledgements: The College Board (2006) AP® Statistics Free-Response Question 5
[Online http://apcentral.collegeboard.com/apc/members/exam/exam_questions/8357.html]
(Used with permission to Nancy Pelaez, npelaez@purdue.edu)

Background Information

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

a) Consider the treatments that the biologist plans to use in this experiment. Create a visual representation and explain the treatments that the biologist plans to use.

Salinity levels: low and high

shrimp

nutrient A   Nutrient B   Nutrient C   (×2 each)

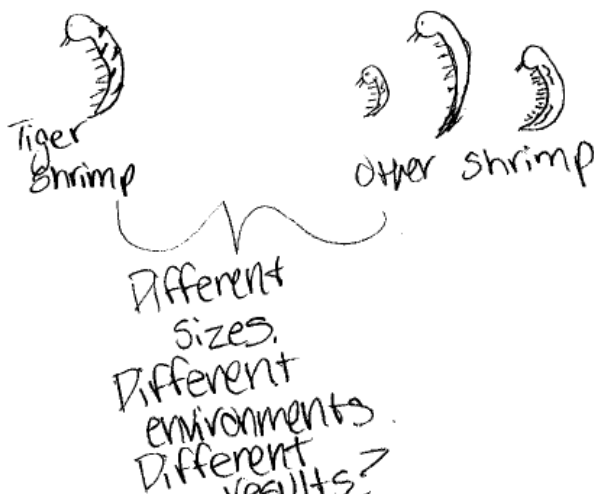The biologists intend to use tiger shrimp to test 3 different growth-enhancing nutrients and 2 salinity levels

b) Using the treatments listed in part (a); describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks. What visual representation will allow the biologist to compare the shrimps' growth after 3 weeks?

In order to compare the growth, the average size of the shrimp must be known. This means that there must be a "control" to compare the experimental results to. The control's results should be recorded and noted before examining the results of the other experimental results. With the average size recorded, the results of the other tanks have a basic unit for comparison. Remove the shrimp from each of the 12 tanks and record their growth [Be sure that prior to the start of the experiment the shrimps' original sizes must be recorded] once the new growth sizes have been recorded, look back to their original sizes. How do they compare? Then compare each growth result to the control. Is there a significant difference in growth? Are the control shrimp larger or smaller than the control? (The control representation.)

1

c) Consider one statistical advantage to having only tiger shrimps in the experiment. To explain why this is an advantage, create a visual representation that illustrates one statistical advantage to having only tiger shrimps in the experiment. If only tiger shrimp are used, it will be easier to determine a corollation between the growth of the shrimp and the salinity level. For example, it would not be accurate if various types of shrimp were used, as different species can be affected in significantly different ways.
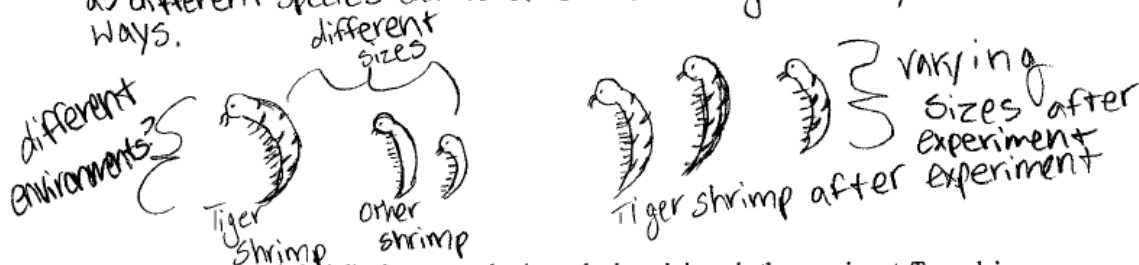
d) Consider one statistical disadvantage to having only tiger shrimps in the experiment. To explain why this is a disadvantage, create a visual representation that illustrates one statistical disadvantage to having only tiger shrimps in the experiment. If only tiger shrimp are used, the corollation, or lack of, between growth and salinity level will only be accurate for this one species. Other shrimp could react differently in this same experiment. So the results would only reflect how these nutrients and salinity level after tiger shrimp.


2

Question 2: The "Drug" Question

Background Information

The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing.

Directions

As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

a) Using complete sentences state the hypothesis to be tested. Create a visual representation to explain your hypothesis. *If this drug were to be tested on humans, then the result would be positive and harmless.*
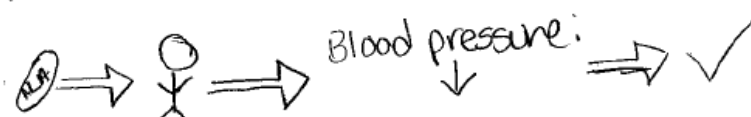
b) Since there are several contributing factors that can affect blood pressure levels, list **five** factors that will be constant between the experimental and control groups. Create a visual representation that illustrates these constant factors between experimental and control group. *There are many factors that can affect blood pressure. In order to make the experiment more accurate, we must keep the following items constant: gender, race, age range, activity level and medical history*

c) Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study. What visual representation will the scientist use to explain your answer? *Since many different factors affect blood pressure, certain criteria is needed in choosing who participates. For example, if we use a variety of races, the results may differ greatly. This will make the experiment less accurate. If the criteria is not set, the results will be different and the overall affect of the drug could be unclear. Utilize the ethnic*

1

d) Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group. Create a visual representation to explain how participants will be selected as a member of the experimental or control group. In order to make the selection more "natural", the selections should be done at random. From there, determine which will be in the control or experimental group. Seperate draws can be done for both groups.
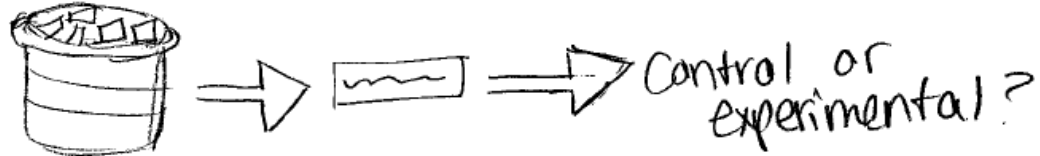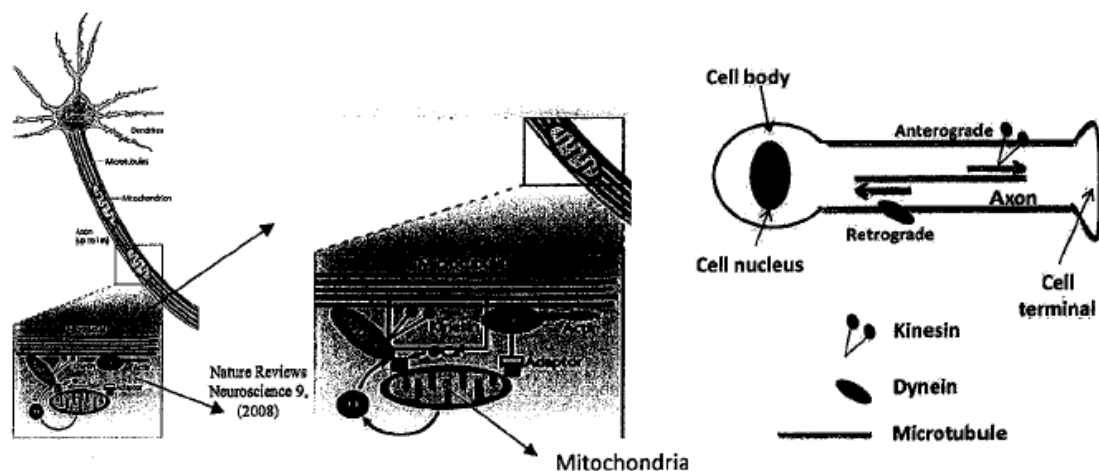

e) Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken. What visual representation will allow the scientist to make measurements to judge the efficiency of Alamain? In order to judge the drug's efficiency, the blood pressure of the participates should be recorded frequently. How frequently? Perhaps daily or weekly. This experiment should take place of several months, to ensure the drug has been given enough time to work through the system. After the experiment is complete, create a line graph of each participates blood pressure over the course of the experiment.

f) Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans. What visual representation will allow the scientist to indicate the success or failure of the drug Alamain? Determine the average blood pressure rating for the group. How do the results compare to the average? Maybe create a bar graph for each participants average over the entire experiment. How do the new averages compare to the overall average? If there is a positive difference, meaning the new average is significantly lower than the orignal average, then the drug could be declared "successful".

Question 3: The "Neuron" Question

Acknowledgements: Original question by Annwesa Dasgupta

Figures

Nature Reviews Neuroscience 9, (2008)

Mitochondria

Background

Mitochondria are one of the several organelles that get transported across the axon of a nerve *(Refer figure above)*. They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions

Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder caused by impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons.

The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;

Compound D: inhibits dynein;

An Image software: measures mitochondrial movement in neurons.

- How do you think a 'hypothesis' relates to an experiment? Create a visual representation to illustrate your answer. The hypothesis gives you a basic idea to test for the purpose of the experiment.

  Hypothesis → experiment → proved or disproved?

a) Describe what you see in the three diagrams above. Please tell us in detail what you think about it. The three diagrams provide a visual representation of the types of axon transports. They all show the basic layout of the cell and where the transport processes occur. Each diagram shows where the kinesin and dynein are located with respect to the cell. The diagrams are elaborate and seem to portray the transports clearly.

b) What could be a potential hypothesis for your experiment? Create a visual representation to illustrate your hypothesis. If the compounds are inhibited or prohibited, then there should be a visable effect on the mitochondrial movement in neurons.

△ kinesin
☐ dynein
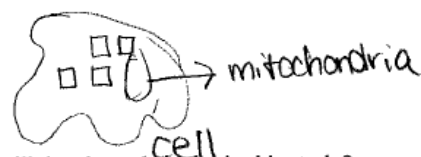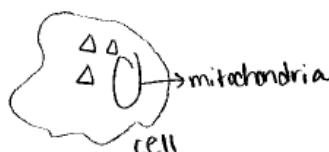
c) Which factors will the researchers vary and which will they keep the same in this study? Why? Use a visual representation to explain the factors they will vary and keep the same. The compounds will generally remain the same. However, the amounts of the compounds should differ. The types of cell used must also be the same and the time period over which this experiment takes place should also remain the same.

Amount kinesin/dynein   Time   Cell

d) Consider how researchers will assign subjects to groups for this experimental study. To explain how they will assign subjects, create a visual representation to support your answer.

healthy cell

Control

cell with disorder

Experiment

e) Consider a cause-and-effect relationship between the treatment and a response variable in this experiment. To justify this relationship, create a visual representation to explain a cause and effect relationship between the treatment and a response variable in the experiment. Does the treatment inhibit transports in the cells with a disorder?

cell + △ kinesin and/or □ dynein ⟹ increase in mitochondrial transport
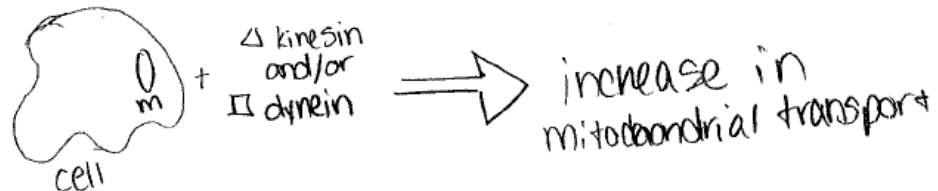
f) How would the researchers present the results of this experiment?
They would create a graph showing the mitochondrial movements in the neurons. How much movement occured?

g) What results do the researchers expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of this experiment. What visual representation will allow them to present results?
Researchers should expect to see a difference in the movements with regard to the kinesin and dynein. Areas of little movement mean that a vitable portion of the process was not present. Compare areas of considerable movement to areas of little to no movement. If the healthy amount of movement occurs, then it was successful. If there was little to no movement, then it failed. Utilize graphs to help determine the results

3

h) Consider improving the validity of the experiment. What visual presentation will you use to show how will the validity be improved? In order to determine if the experiment is valid, the "healthy" amount of mitochondrial movements should be known. Use bar graphs to compare the validity in the experiments

i) What do you think this diagram is not showing? Explain your answer.
The diagram does not show exactly which motor protein is in charge of the movements. Are both needed? If not, which is responsible for both types of movement?

j) Is there anything about this question that you don't understand or find confusing? Explain. Part E was a bit confusing to me. This could be due to my minimal knowledge of this subject or my lack of understanding how this process works. Perhaps more info would be needed for me to better understand how to answer that part?

k) Consider yourself a diagram designer. If you could change the diagrams, what would you change or how would you improve them? The diagrams should specifically show which motor protein is responsible for anterograde and retrograde transport. Show if only one is needed or if both are needed for the transports to occur

4

Li Na Written Assessment Response

The questions on this page do not require much scientific knowledge. Most scientists can answer these questions. Please write down your own ideas in response to these questions. There is no time or word limit so feel free to share your thinking as much as you can. Please ask for more writing paper if you need.

Question 1: The "Shrimp" Question

Acknowledgements: The College Board (2006) AP® Statistics Free-Response Question 5
[Online http://apcentral.collegeboard.com/apc/members/exam/exam_questions/8357.html]
(Used with permission to Nancy Pelaez, npelaez@purdue.edu)

Background Information

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

a) Consider the treatments that the biologist plans to use in this experiment. Create a visual representation and explain the treatments that the biologist plans to use.

Treatments: Biologists treat the shrimps with different salinity in water and growth-enhancing nutrients to see the effects on growth of shrimps

A + low → group 1 and 2
A + high → group 3 and 4
B + low → group 5 and 6
B + high → group 7 and 8
C + low → group 9 and 10
C + high → group 11 and 12

Biogists have two variables, the nutrients and salinity. So make each 2 tanks treated with one combination (nutrients + salinity). The replication in one part avoids the chance variables + make the experiment more trustful

b) Using the treatments listed in part (a); describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks. What visual representation will allow the biologist to compare the shrimps' growth after 3 weeks?

The 120 tiger shrimps should be similar from the same waters (the original environments) healthy. same body length, same age (larvae have the potential to grow), gender (just female/male because different genders may react differently

Assign all the shrimps into 12 groups without any pre-classification After 3 weeks biologists can compare the body length of the shrimps to see how much they have grown during those 3 weeks.

1

→ Cannot tell the difference

c) Consider one statistical advantage to having only tiger shrimps in the experiment. To explain why this is an advantage, create a visual representation that illustrates one statistical advantage to having only tiger shrimps in the experiment.

Other species may cause confounding. The causation may be misled for different species may have different growth speed to the same body length due to the hormone and gene. [the morphogen in embryo degradades to different position making the body length varying]

The advantage is to control all the variables only constrained into the different growth-enhancing nutrients and salinity levels. Though other species may react similarly but the growth during the same time cannot give a more apparent tendancy to find a stronger causation.

d) Consider one statistical disadvantage to having only tiger shrimps in the experiment. To explain why this is a disadvantage, create a visual representation that illustrates one statistical disadvantage to having only tiger shrimps in the experiment.

For biologists want to find the effects on shrimps. Only tiger shrimps can not be the representative specie of all the shrimps. They may be an exception to the salinity and different nutrients, and they may be polyphelitic with other species. Even though they have converge traits but they don't share the common ancestor (different genes, hormones) which make them not representative of all the species.

Strong tendancy.

Question 2: The "Drug" Question

Background Information

The drug ALAMAIN has been developed by the Gentronic Drug Company to (lower blood pressure) in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing.

Directions

As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the (new high blood pressure drug) Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

a) Using complete sentences state the hypothesis to be tested. Create a visual representation to explain your hypothesis.

Alamain can lower the human blood pressure
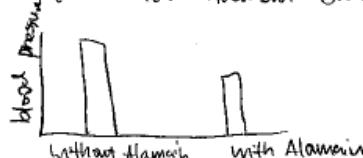
b) Since there are several contributing factors that can affect blood pressure levels, list **five** factors that will be constant between the experimental and control groups. Create a visual representation that illustrates these constant factors between experimental and control group.

① Age

② gender

③ nationality (and species American/European/Asian)

④ similar living habit

⑤ similar body condition (similar blood pressure, body build)

c) Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study. What visual representation will the scientist use to explain your answer?

Age may influence the functioning mechanism of human bodies. Males may happen to have higher blood pressure than female. Controling the species may due to the difference of human species (family inherited high blood pressure). Good living habit and good body condition → low blood pressure. So participants should have the constant variables

d) Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group. Create a visual representation to explain how participants will be selected as a member of the experimental or control group,

They may be assinged in block randomly experiment. They can be divided into different groups with different age range, gender, health conditioning. Control group should be only one factor changing to make the comparisons between the experimental and control groups more apparent.

e) Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken. What visual representation will allow the scientist to make measurements to judge the efficiency of Alamain?

Scientist will measure the blood pressure of the participants within the efficient time (the same) they take the Alamain. They should take the measurement every thirty minutes (or shorter) to see the efficiency of the Alamain
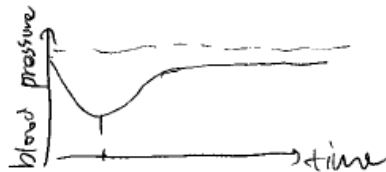
f) Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans. What visual representation will allow the scientist to indicate the success or failure of the drug Alamain?

Referring the graph at e. In an efficient time range the blood pressure can go down relatively but they may bounce back in a longer time because Alamain fails to keep the blood pressure constantly low if it is not continually taken

Question 3: The "Neuron" Question

Acknowledgements: Original question by Annwesa Dasgupta

Figures

Nature Reviews Neuroscience 9, (2008)

Cell body
Cell nucleus
Anterograde
Retrograde
Axon
Cell terminal
Kinesin
Dynein
Microtubule
Mitochondria

Background

Mitochondria are one of the several organelles that get transported across the axon of a nerve (*Refer figure above*). They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions

Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder caused by impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons.

The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;

Compound D: inhibits dynein;

An Image software: measures mitochondrial movement in neurons.

- How do you think a 'hypothesis' relates to an experiment? Create a visual representation to illustrate your answer.

    Kinesin can cause mitochondria to anterograde while dynein can cause the mitochondria to retrograde. Both of them activates can cause the disorder.

a) Describe what you see in the three diagrams above. Please tell us in detail what you think about it.

    While Mitochondria is the place where ATP is produced in a large amount and myosin is the acting protein which can push the mitochondria to move. The kinese and dynein have different ion (Ca²⁺ or Na⁺?) pump channels to the different direction to move the mitochondria in the different directions.

b) What could be a potential hypothesis for your experiment? Create a visual representation to illustrate your hypothesis.

    The cell terminal has more negative charges than the cell body.
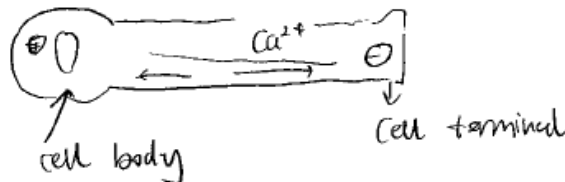
    cell body   Ca²⁺   cell terminal

c) Which factors will the researchers vary and which will they keep the same in this study? Why? Use a visual representation to explain the factors they will vary and keep the same.

    The same animal cell with the same mitochondria, myosine the same amount of ATP in the similar environment). The concentrations of the external cellular complexes should be same the maintain the same diffusion pressure.

d) Consider how researchers will assign subjects to groups for this experimental study. To explain how they will assign subjects, create a visual representation to support your answer.

① inhibits kinesin + activates Dynein
② inhibits kinesin + inhibits Dynein
③ activates kinesin + inhibits Dynein
④ activates kinesin + activates Dynein
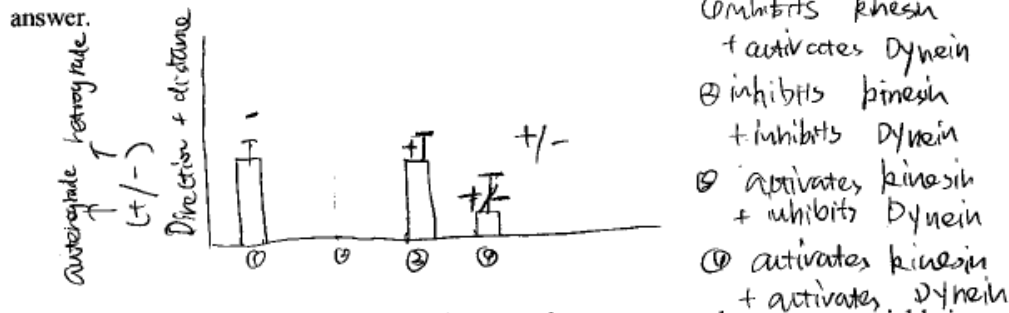
e) Consider a cause-and-effect relationship between the treatment and a response variable in this experiment. To justify this relationship, create a visual representation to explain a cause and effect relationship between the treatment and a response variable in the experiment.

Kinesin causes the mitochondria to do the anterograde transporting while dynein causes the mitochondria to do the retrograde transporting due to the ions interactions.
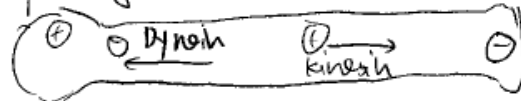
f) How would the researchers present the results of this experiment?

They can record the image of how mitochondria move. (displacement/time) anterograde + direction and retrograde - direction.

g) What results do the researchers expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of this experiment. What visual representation will allow them to present results?

Researchers may find activates kinesin and dynein can affect (successfully binding) the direction the mitochondria moves. But different concentration or ATP amount required make them no difference to tell the ②, ④. Both activates kinesin and activates dynein may cancel each other and may have the same response as the group ②

3

h) Consider improving the validity of the experiment. What visual presentation will you use to show how will the validity be improved?

Improve the group 1,3,4 with different Concentration/ATP amount. In 1,3 set more groups to see the concentration/ATP amount effect on the displacement. And set the relative concentration/ATP amount gradient in group 4 to see the cancel effect of group 4. (Both direction and displacement)

i) What do you think this diagram is not showing? Explain your answer.

It doesn't show the ions in Dynein. The purple dots of kinesin is mainly the Ca^{2+} which binds to the adaptor but dyne doesn't have any ions shown in the graph

j) Is there anything about this question that you don't understand or find confusing? Explain.

How are the dynein/kinesin activated/inhibited at one transportion? Do they require the different messangers? Just randomly. If no disorder, how can the two messangers regulate each other?

k) Consider yourself a diagram designer. If you could change the diagrams, what would you change or how would you improve them?

Daniel Written Assessment Response

The questions on this page do not require much scientific knowledge. Most scientists can answer these questions. Please write down your own ideas in response to these questions. There is no time or word limit so feel free to share your thinking as much as you can. Please ask for more writing paper if you need.

Question 1: The "Shrimp" Question

Acknowledgements: The College Board (2006) AP® Statistics Free-Response Question 5
[Online http://apcentral.collegeboard.com/apc/members/exam/exam_questions/8357.html]
(Used with permission to Nancy Pelaez, npelaez@purdue.edu)
Background Information

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A. B. and C) and two different salinity levels (low and high).

a) Consider the treatments that the biologist plans to use in this experiment. Create a visual representation and explain the treatments that the biologist plans to use.

There will be a control group, along with 5 of the tanks which will only have 1 variable different from the control, and then 6 tanks which will combine a nutrient and salinity level to determine how each

b) Using the treatments listed in part (a); describe a completely randomized design that will allow affects grew the biologist to compare the shrimps' growth after 3 weeks. What visual representation will allow the biologist to compare the shrimps' growth after 3 weeks?

First randomly assign shrimp to each tank, measuring them beforehand. Then apply the treatments shown above and measure them again. To determine growth not only should length be measured but also mass and may also be measured if necessary. A comparison between tanks can then be made to determine which produced the most growth.

c) Consider one statistical advantage to having only tiger shrimps in the experiment. To explain why this is an advantage, create a visual representation that illustrates one statistical advantage to having only tiger shrimps in the experiment.

By only using tiger shrimps, the researcher will not have to take into account the differing growth rates of different shrimp as well as the fact that different shrimp may grow better under different conditions.

d) Consider one statistical disadvantage to having only tiger shrimps in the experiment. To explain why this is a disadvantage, create a visual representation that illustrates one statistical disadvantage to having only tiger shrimps in the experiment.

By only using tiger shrimp, the results of the experiment may not be able to be applied to other shrimp because as stated above, different shrimp may grow better under different conditions, meaning for the results to apply to all shrimp a study on each species must be done.

Question 2: The "Drug" Question

Background Information

The drug ALAMAIN has been developed by the Gentronic Drug Company to lower blood pressure in people whose blood pressure is too high. The drug has been thoroughly tested on animals with positive results. The Gentronic Drug Company feels it is now time for the drug to be tested on humans, and have contacted the Human Improvement Laboratory (HIL) to do the testing.

Directions

As chief research scientist at the Human Improvement Laboratory (HIL) you have been assigned the task of developing the human testing program for the new high blood pressure drug Alamain. You and your assistants are to confer on the experimental design of this testing program, and to write a report outlining the program. The report is to be submitted to the chairperson of the HIL Drug-Testing Committee for approval. Complete the following sections as you would include them on your report.

a) Using complete sentences state the hypothesis to be tested. Create a visual representation to explain your hypothesis.

IF Alamain is given to patients, then there blood pressure will become lower.

b) Since there are several contributing factors that can affect blood pressure levels, list **five** factors that will be constant between the experimental and control groups. Create a visual representation that illustrates these constant factors between experimental and control group.

1) Doctor

2) Stress-level (relative to others)

3) Diet
4) Body weight
5) Sex (have 2 control groups 1 male, 1 female)

c) Based on the factors listed in Question 2, using complete sentences explain why certain criteria need to be used in choosing the participant in this study. What visual representation will the scientist use to explain your answer?

By using the same doctor, each patient will be receiving the same treatment. A survey should be given to measure stress level so if amount of stress lowers during treatment or gets higher it can be accounted for. For Diet and body weight it may affect the groups differently due to their metabolism rate - and it may have differing effects on men and women.

1

d) Once the list of the participants has been created, using complete sentences explain how they will be selected to be a member of either the experimental or control group. Create a visual representation to explain how participants will be selected as a member of the experimental or control group.

First each patient will be numbered, then using a random number generator patients will be put into there groups. The 1st 50% chosen will go into experimental and the other 50% go into control. If more than 1 control or experimental group is needed such as splitting into male/female then number all males 1-? and females 1-? and put them into groups as stated above, giving 4 total rather than 2 groups.

e) Using complete sentences, explain what measurements and/or tests will be made on the experimental and control groups to judge the efficiency of Alamain, and how often measurements or test will be taken. What visual representation will allow the scientist to make measurements to judge the efficiency of Alamain?

Blood pressure will be measured as well as diet and Stress level as well as other variables which may affect the results. At least once every 1 or 2 weeks depending on the length of the experiment and the number of patients a doctor is in charge of.

may be different should be determined by doctor

f) Using complete sentences, explain what criteria will be used to indicate the success or failure of the drug Alamain to reduce blood pressure levels in humans. What visual representation will allow the scientist to indicate the success or failure of the drug Alamain?

If those in the experimental group have a larger drop in blood pressure (on average) than those in the control group then Alamain was successful, otherwise it fails,

Question 3: The "Neuron" Question

Acknowledgements: Original question by Annwesa Dasgupta

Figures

Nature Reviews Neuroscience 9, (2008)

Mitochondria

Cell body
Anterograde
Axon
Retrograde
Cell nucleus
Cell terminal

Kinesin
Dynein
Microtubule

Background

Mitochondria are one of the several organelles that get transported across the axon of a nerve *(Refer figure above)*. They are transported in both directions along the length of the axon. The movement of mitochondria from the cell body to the cell terminal is termed as anterograde transport while the movement from the cell terminal to the cell body, in the opposite direction, is termed as retrograde transport. Movement of mitochondria takes place on the microtubules present along the length of the axons. This complex movement is facilitated by the interaction of motor proteins, kinesin and dynein, present in the axons.

Directions

Medical researchers at Seattle Grace Hospital are trying to diagnose the cause for a disorder caused by impaired mitochondrial movement within neurons in human subjects. Cell culture studies have been performed to observe the movement of mitochondria within neurons.

The researchers think that kinesin or dynein activity might play a role in the cause of this disorder. Pretend that you work for a company called *MedResearch* that has been assigned to design an experiment to test how kinesin or dynein can effect mitochondrial movement. In your lab you have the following chemicals:

Compound K: inhibits kinesin;

Compound D: inhibits dynein;

An Image software: measures mitochondrial movement in neurons.

- How do you think a 'hypothesis' relates to an experiment? Create a visual representation to illustrate your answer.

  A hypothesis is an educated guess as to how the results of the experiment will come out.

a) Describe what you see in the three diagrams above. Please tell us in detail what you think about it. The first diagram, shows an "up-close" view of a mitochondrial cell in the axon of a nerve. The second diagram is the same as the first except it emphasizes what the mitochondria is doing within the axon. The 3rd diagram gives a basic overview of the process showing that Dynein moves toward the terminal body while Kinesin moves away from it.

b) What could be a potential hypothesis for your experiment? Create a visual representation to illustrate your hypothesis.

  If dynein is inhibited, then mitochondrial movement toward the terminal body will stop and if kinesin is inhibited then mitochondrial movement away from the terminal body will stop.

c) Which factors will the researchers vary and which will they keep the same in this study? Why? Use a visual representation to explain the factors they will vary and keep the same.

  They must keep the same type of axon as well as the conditions the mitochondria are held under. They will then vary the treatment given to determine the change in mitochondria movement, if any.

d) Consider how researchers will assign subjects to groups for this experimental study. To explain how they will assign subjects, create a visual representation to support your answer. They will be randomly picked t assigned into 3 groups, one give compound K, another given compound D and a 3rd as a control.

e) Consider a cause-and-effect relationship between the treatment and a response variable in this experiment. To justify this relationship, create a visual representation to explain a cause and effect relationship between the treatment and a response variable in the experiment. As each treatment is given, it may stop or slow mitochondrial movement which can then be measured with the image software to determine any effects the treatment had on movement.

f) How would the researchers present the results of this experiment? They would present the movement seen in the control group (on average) and then compare it to the movement seen in the other 2 groups.

g) What results do the researchers expect to get and what would those mean? Using complete sentences, explain what criteria will be used to indicate the success or failure of this experiment. What visual representation will allow them to present results? They expect to see movement stop if the only proteins controlling movement are Dynein and Kinesin

h) Consider improving the validity of the experiment. What visual presentation will you use to show how will the validity be improved?

Multiple experiments could be done on differing axons, other variables could also be changed to see if they effected mitochondrial movement.

i) What do you think this diagram is not showing? Explain your answer.

It does not show the relationship between the 2 proteins, diagrams 1 + 2 appear to show that kinesin relies on dynein to be produced while 3 does not leading to uncertainty of their relationship.

j) Is there anything about this question that you don't understand or find confusing? Explain.

Diagram 2 is confusing because I don't fully understand the process that is being shown

k) Consider yourself a diagram designer. If you could change the diagrams, what would you change or how would you improve them?

Make the diagrams less detailed if possible, leaving out unnecessary pieces of information,

APPENDIX M: Participant Perception Inventory (PPI)

The next items are designed to measure your perception of your knowledge, experience, and confidence on various topics that will be covered in this course. Indicate how you feel about your knowledge, experience, and confidence
(where a great deal = 4   Average = 3   None = 1).

EXAMPLE:	**knowledge**	**experience**	**confidence**
Changing a flat tire.	⑤ 4 3 2 1	5 4 ③ 2 1	5 4 3 2 ①

This would mean that I have a great deal of knowledge about changing a flat tire (response of 5), I have an average amount of experience with changing a flat tire (response of 3), but I am not confident (response of 1) in my ability to change a flat tire.

Indicate your feelings of knowledge, experience, and confidence about the following:

A great deal = 5   Much = 4   Average = 3   A little = 2   None = 1

A. Physical and Chemical Basis of Life
a1. Understanding how acid-base equilibria (pH and buffers) influence partitioning of molecules in body compartments.

1. knowledge
2. experience
3. confidence

a2. Understanding the size and structure of second messengers such as calcium, cyclic AMP, IP3 and DAG.

4. knowledge
5. experience
6. confidence

a3. Explaining what kinds of bonds fold proteins and nucleic acids into a three-dimensional shape.

7. knowledge
8. experience
9. confidence

a4. Explaining how a protein kinase as part of protein signaling network can alter protein-protein interactions.

10.   knowledge
11.   experience

12.   confidence

a5. Using appropriate representations to draw biological molecules and macromolecules.

13.   knowledge
14.   experience
15.   confidence

B. **Molecular Basis of Regulation**
b1. Understanding how membrane potentials are generated and describing this process mathematically using the Nernst equation.

16.   knowledge
17.   experience
18.   confidence

b2. Distinguishing properties of excitable from non-excitable cells within different systems of the body.

19.   knowledge
20.   experience
21.   confidence

b3. Explaining examples of responses regulated by G-protein coupled receptors.

22.   knowledge
23.   experience
24.   confidence

b4. Explaining mechanisms by which different cells use neurotransmitters to communicate and coordinate.

25.   knowledge
26.   experience
27.   confidence

b5. Explaining the specificity for control by biological signal transduction pathways.

28.   knowledge
29.   experience
30.   confidence

C. **Plant Biology**
c1. Understanding how plants regulate their own water handling.

31. knowledge
32. experience
33. confidence

c2. Understanding how hormones such as auxins regulate plant growth.

34. knowledge
35. experience
36. confidence

c3. Understanding how calcium influx upon gamete fusion prevents polyspermy before egg activation and the initiation of development in plants.

37. knowledge
38. experience
39. confidence

c4. Identifying mechanisms an organism can use to control osmotic pressure.

40. knowledge
41. experience
42. confidence

c5. Comparing the mechanisms and the outcomes of self-pollination and fertilization in flowering plants.

43. knowledge
44. experience
45. confidence

D. Animal Biology

d1. Recognizing conditions that alter oxygen handling in mammals.

46. knowledge
47. experience
48. confidence

d2. Understanding signals involved in the shaping of animal body plans in development and evolution.

49. knowledge
50. experience
51. confidence

d3. Understanding what causes shortening or force development of a muscle.

52.   knowledge
53.   experience
54.   confidence

d4. Understanding how the heart and blood vessels regulate transport in the human body.

55.   knowledge
56.   experience
57.   confidence

d5. Explaining how apical and basolateral membranes function to transport substances across epithelial cell layers.

58.   knowledge
59.   experience
60.   confidence

E. Experimental Biology
e1. Identifying whether data is quantitative or categorical.

61.   knowledge
62.   experience
63.   confidence

e2. Identifying whether an investigation uses observation or an experimental approach.

64.   knowledge
65.   experience
66.   confidence

e3. Choosing the best way to graphically represent data with a histogram, scatterplot, time course graph, bar chart, dot plot, or side-by-side graph.

67.   knowledge
68.   experience
69.   confidence

e4. Distinguishing causality from correlation based on association between variables.

70.   knowledge
71.   experience
72.   confidence

e5. Describing a carefully controlled experiment from a biological research paper.

73. knowledge
74. experience
75. confidence

F. Biological Information Literacy

f1. Defining a research question related to the unity and the diversity of life and how organisms work.

76. knowledge
77. experience
78. confidence

f2. Reading primary literature, scientific web resources, and research reviews to find out about an investigation that illustrates how organisms work.

79. knowledge
80. experience
81. confidence

f3. Locating, identifying, and retrieving information resources to learn about how organisms work.

82. knowledge
83. experience
84. confidence

f4. Evaluating and treating critically information about how and why knowledge has changed in biology.

85. knowledge
86. experience
87. confidence

f5. Citing scientific research sources and using the information ethically and legally in writing about the unity and the diversity of life.

88. knowledge
89. experience
90. confidence

Appendix N: Descriptive Statistical Analysis Tests for the PPI Assessment

Average KEC for learning outcome categories and underlying statements arranged in increasing order of category means in an introductory level biology course

Categories and Statements	Pre				Post				Effect sizes
	Category Means	SD	95% confidence	Cronbach's α	Category Means	SD	95% confidence	Cronbach's α	
Molecular Basis of Regulation	**2.02**	--	--		**3.61**	--	--		--
G-protein coupled receptors	1.76	0.91	0.12		3.69	0.93	0.12		0.72**
Membrane potential and Nernst equation	1.91	1.07	0.14		3.62	0.96	0.12		0.64**
Excitable vs. non-excitable cells	1.98	0.99	0.13	0.95	3.54	0.90	0.12	0.95	0.64**
Why signal transduction exists	2.14	1.00	0.13		3.56	0.90	0.12		0.60**
Communication using neurotransmitters	2.33	1.06	0.14		3.65	0.87	0.11		0.56*
Plant Biology	**2.37**	--	--		**3.62**	--	--		--
Calcium influx in gamete fusion	1.77	0.94	0.12		3.61	0.93	0.12		0.70**
Plant hormone regulation	1.96	1.05	0.13		3.85	0.88	0.11		0.70**
Plant osmotic pressure	2.55	1.16	0.15	0.95	3.48	0.86	0.11	0.95	0.41*
Plant water handling	2.68	1.12	0.14		3.71	0.86	0.11		0.46*
Plant self pollination and fertilization	2.86	1.12	0.14		3.45	0.87	0.11		0.28
Animal Biology	**2.37**	--	--		**3.60**	--	--		--
Transportation across cell layers	1.87	0.98	0.13	0.95	3.21	0.94	0.12	0.94	0.57*
Signaling in animal body	2.25	1.07	0.14		3.59	0.80	0.10		0.58*

Average KEC for learning outcome categories and underlying statements arranged in increasing order of category means in an introductory level biology course

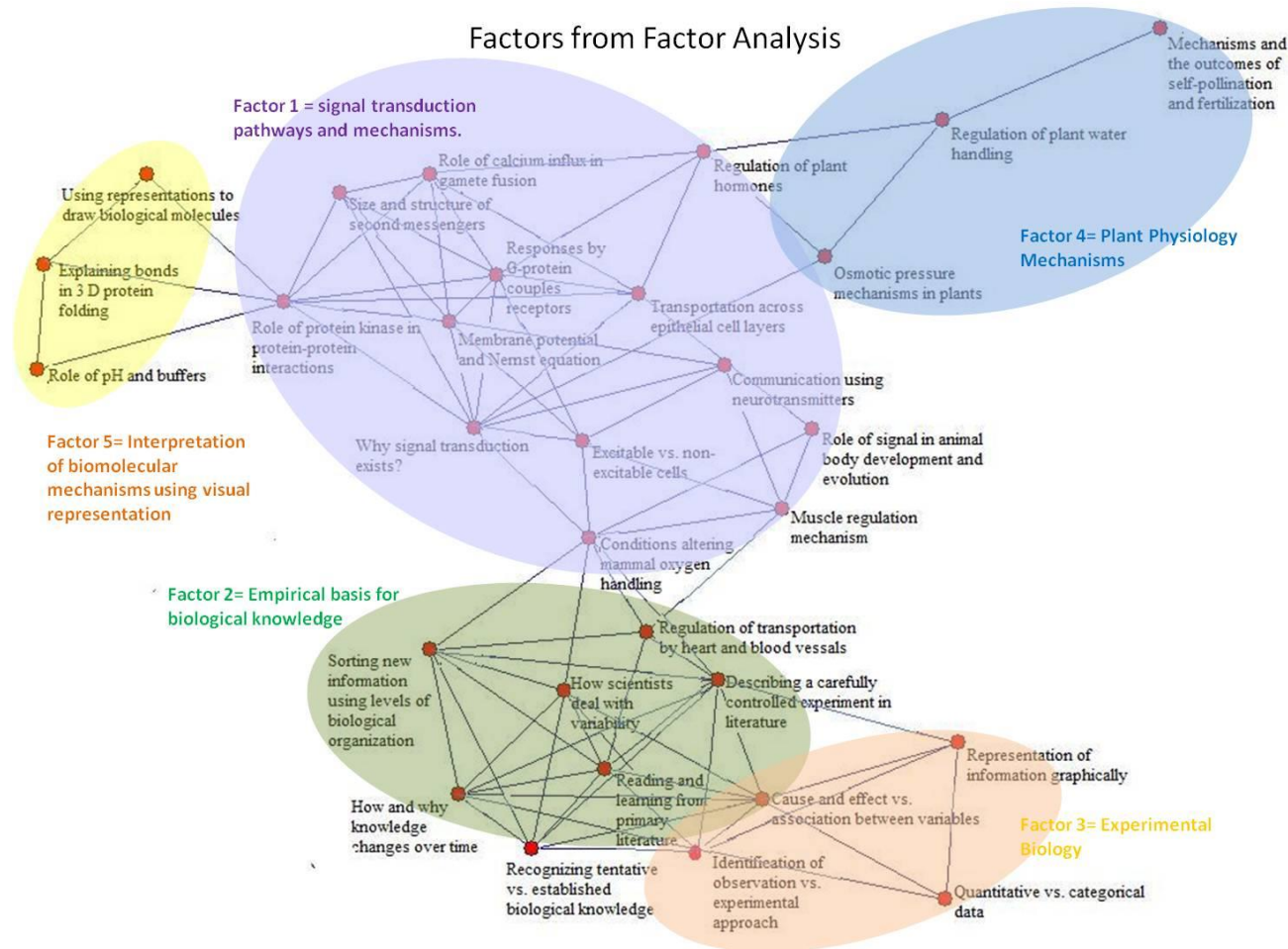| Categories and Statements	Pre				Post				Effect sizes
	Category Means	SD	95% confidence	Cronbach's α	Category Means	SD	95% confidence	Cronbach's α	
development and evolution									
Muscle regulation	2.28	1.18	0.15		3.74	0.92	0.12		0.57*
Mammal oxygen handling	2.41	1.08	0.14		3.54	0.78	0.10		0.51*
Transportation by heart and blood vessels	3.05	1.14	0.15		3.89	0.81	0.10		0.39*
Physical and Chemical Basis of Life	**2.57**	--	--		**3.52**	--	--		--
Size and structure of second messengers	1.83	0.97	0.12		3.44	0.97	0.12		0.64**
Role of protein kinase	2.22	0.99	0.13	0.94	3.48	0.92	0.12	0.95	0.55*
Role of pH and buffers	2.64	0.96	0.12		3.47	0.96	0.12		0.40*
Explain 3D protein folding bonds	3.06	1.11	0.14		3.77	1.01	0.13		0.32*
Visual representation of biological molecules	3.10	1.11	0.14		3.44	0.89	0.11		0.17
Empirical basis of biological knowledge	**2.89**	--	--		**3.76**	--	--		--
Dealing with variability	2.65	1.08	0.14		3.58	0.75	0.10		0.45*
Sorting information with biological organization levels	2.71	1.11	0.14	0.97	3.51	0.86	0.11	0.94	0.37*
Tentative vs. Established biology	2.82	1.14	0.15		3.68	0.85	0.11		0.39*
Learning from primary literature	2.97	1.23	0.16		3.82	0.85	0.11		0.37*

Average KEC for learning outcome categories and underlying statements arranged in increasing order of category means in an introductory level biology course

Categories and Statements	Pre				Post				Effect sizes
	Category Means	SD	95% confidence	Cronbach's α	Category Means	SD	95% confidence	Cronbach's α	
How/why knowledge changes over time.	3.30	1.16	0.15		4.20	0.74	0.10		0.42*
Experimental Design	**3.25**	--	--		**4.14**	--	--		--
Experiment description from literature	2.95	1.21	0.15		3.95	0.75	0.10		0.44*
Quantitative vs. Categorical Data	3.20	1.20	0.15	0.97	4.33	0.75	0.10	0.94	0.49*
Distinguishing causality from correlation	3.29	1.20	0.15		3.91	0.78	0.10		0.29
Graphical representation	3.37	1.17	0.15		4.16	0.73	0.09		0.38*
Observation vs. Experimental approach	3.41	1.20	0.15		4.36	0.70	0.09		0.44*

Appendix O: Clusters of Biology Knowledge Areas from Factor Analysis

Factors from Factor Analysis

Factor 1 = signal transduction pathways and mechanisms.

Factor 4= Plant Physiology Mechanisms

Factor 5= Interpretation of biomolecular mechanisms using visual representation

Factor 2= Empirical basis for biological knowledge

Factor 3= Experimental Biology

Mechanisms and the outcomes of self-pollination and fertilization

Regulation of plant water handling

Role of calcium influx in gamete fusion

Regulation of plant hormones

Using representations to draw biological molecules

Size and structure of second messengers

Responses by G-protein couples receptors

Osmotic pressure mechanisms in plants

Explaining bonds in 3 D protein folding

Transportation across epithelial cell layers

Role of protein kinase in protein-protein interactions

Membrane potential and Nernst equation

Communication using neurotransmitters

Role of pH and buffers

Role of signal in animal body development and evolution

Why signal transduction exists?

Excitable vs. non-excitable cells

Muscle regulation mechanism

Conditions altering mammal oxygen handling

Regulation of transportation by heart and blood vessals

Sorting new information using levels of biological organization

How scientists deal with variability

Describing a carefully controlled experiment in literature

Representation of information graphically

Reading and learning from primary literature

Cause and effect vs. association between variables

How and why knowledge changes over time

Recognizing tentative vs. established biological knowledge

Identification of observation vs. experimental approach

Quantitative vs. categorical data

Pre-instruction network analysis to visualize the factors represented here in color codes. 'Plant Physiology' items cluster separately from 'Animal Biology' items (under 'Signal Transduction'). 'Experimental Reasoning' clusters separately from all categories'.

Appendix P: Institutional Review Board Consent Letter

PURDUE
UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	NANCY PELAEZ
	LILY
From:	RICHARD MATTES, Chair
	Social Science IRB
Date:	09/15/2010
Committee Action:	**Expedited Approval**
IRB Action Date:	08/27/2010
IRB Protocol #:	1008009581
Study Title:	How do Students Think about Experimental Design in Biology?
Expiration Date:	09/13/2011

Following review by the Institutional Review Board (IRB), the above referenced protocol has been approved. This approval permits you to recruit subjects up to the number indicated on the application form and to conduct the research as it is approved. The IRB-stamped and dated consent, assent, and/or information form(s) approved for this protocol are enclosed. Please make copies from these document(s) both for subjects to sign should they choose to enroll in your study and for subjects to keep for their records. Information forms should not be signed. Researchers should keep all consent/assent forms for a period no less than three (3) years following closure of the protocol.(6) (7) .

Revisions/Amendments: If you wish to change any aspect of this study, please submit the requested changes to the IRB using the appropriate form. IRB approval must be obtained before implementing any changes unless the change is to remove an immediate hazard to subjects in which case the IRB should be immediately informed following the change.

Continuing Review: It is the Principal Investigator's responsibility to obtain continuing review and approval for this protocol prior to the expiration date noted above. Please allow sufficient time for continued review and approval. No research activity of any sort may continue beyond the expiration date. Failure to receive approval for continuation before the expiration date will result in the approval's expiration on the expiration date. Data collected following the expiration date is unapproved research and cannot be used for research purposes including reporting or publishing as research data.

Unanticipated Problems/Adverse Events: Researchers must report unanticipated problems and/or adverse events to the IRB. If the problem/adverse event is serious, or is expected but occurs with unexpected severity or frequency, or the problem/even is unanticipated, it must be reported to the IRB within 48 hours of learning of the event and a written report submitted within five (5) business days. All other problems/events should be reported at the time of Continuing Review.

We wish you good luck with your work. Please retain copy of this letter for your records.

VITA

VITA

ANNWESA DASGUPTA
425, South River Road**.** Apt – 26**.** West Lafayette**.** IN-47906**.** USA
Ph: 765-430-5157
E-mail: anwesha.dasgupta@gmail.com
Webpage: http://tinyurl.com/Annwesa

EDUCATION

2014   Ph.D.*,* Biological Sciences, Purdue University, USA

Dissertation: Diagnosing Undergraduate Biology Students' Experimental Design Knowledge and Difficulties.

Committee:	Professor Nancy J. Pelaez (advisor)
	Professor David C. Eichinger (chair)
	Professor Trevor R. Anderson
	Professor Dennis J. Minchella
	Professor Jeffrey D. Karpicke

2007   Master of Science, University of Mumbai, India
Thesis: Characterization of Photosynthetic Enzymes from Green Algae

2005   Bachelor of Science, St Xavier's College, Mumbai, India

RESEARCH ACTIVITIES AND PUBLICATIONS

Publications

Dasgupta AP, Anderson TR, Pelaez N (2014). Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties. *CBE Life Science Education 13*, 265-284

Dasgupta AP, Anderson TR, Pelaez N. Design and Development of a 'Neuron Assessment' For Measuring Biology Students' Understanding of Experimental Design. (*In preparation*).

Dasgupta AP, Anderson TR, Pelaez N (2014). Validation of the 'Neuron Assessment' In Comparison To Other Measures of Biology Students' Understanding of Experimental Design. (*In preparation*).

ABSTRACTS AND PRESENTATIONS

- **Dasgupta** AP, Anderson TR, Pelaez N. CUREnet (Course-based undergraduate research experiences), 2014. Invited workshop speaker. *Assessing biology students' learning about experimental design.* Cold Spring Harbor Laboratory (Workshop).
- **Dasgupta** AP, Anderson TR, Pelaez N. *Student Difficulties about Experimental Design in Biology* (2012). EcoLunch Seminar, Department of Biological Sciences, Purdue University (Presentation).
- **Dasgupta** AP, Anderson TR, Pelaez N. Society for Advancement of Biology Education Research (SABER), 2012. Invited Talk. *Designing and Testing of a New Probe to Reveal Student Difficulties with Biological Experiments.* University of Minnesota- Twin Cities (Presentation).
- **Dasgupta** AP, Pelaez N., American Physiological Society, **Research Recognition Award**. *Predicting strategies to improve student perceptions about connections between biology knowledge areas using innovative network analysis software FASEB Journal* 26, 2012 (Poster Abstract).
- Parker LC, **Dasgupta** AP, Adedokun OA, Forney J, Minchella DJ. *A Faculty Learning Community for Integrating Quantitative Statistical Analysis into Undergraduate Biology: Preliminary Impacts and Lessons Learned*. NARST 2012. (Presentation).

TEACHING

- Graduate Teaching Assistant and "super" peer leader for an introductory biology course participating in the Purdue course transformation venture (IMPACT) (Spring 2012). Demonstrated leadership skills in advising teaching interns and facilitating active learning as part of implementation of a cyber Peer-Led Team Learning (cPLTL) program funded by EDUCAUSE (Next Generation Learning Challenges Wave I).
- Instructor for laboratory module in BIOL 20500 - Biology for Elementary School Teachers (Fall 2011 and 2012).
- Graduate Teaching Assistant for a large enrollment introductory biology course (Spring 2010).
- Graduate Teaching Assistant for a large enrollment sophomore level microbiology course (Fall 2007-Spring 2008).

Teaching Awards

- Excellence in Teaching Award 2012-2013, Graduate School, Office of the Provost, Purdue University.
- Outstanding Graduate Teaching Assistant 2012, Biological Sciences, Purdue University.

- Research Recognition Award, 2012, Teaching Section of the American Physiological Society.

PROFESSIONAL DEVELOPMENT EXPERIENCE
- Collaborated on the assessment design and evaluation component for a NSF TUES-type 1 project to report improvements in student skills relevant to fundament ecology and perception changes as a result of participation in ecology project based courses (Fall 2012 -Summer 2013).

- Partnered with biology faculty to design learning modules as part of Faculty learning community component for the HHMI funded project, *Deviating from the Standard: Integrating Statistical Analysis and Experimental Design into Life Science Education.* (Fall 2012-Spring 2013).

- Evaluated and reported accomplishments for the HHMI-funded project: *Deviating from the Standard: Integrating Statistical Analysis and Experimental Design into Life Science Education* at Purdue University (Fall 2010- Spring 2011) (http://hhmi.bio.purdue.edu/).

- Evaluated a Lilly Foundation Inc. endowed undergraduate research internship program and formulated strategies to augment cross disciplinary research experiences for undergraduate students at Purdue (Spring 2010- Fall 2010; Summer 2011) (http://tinyurl.com/DURIPurdue).

ADDITIONAL RESEARCH EXPERIENCE

Bachelors and Masters Program
- Experience with laboratory procedures like with fundamental cell and molecular biology techniques like genomic DNA isolation from bacterial and viral cultures, bacterial plasmid extractions, genetic karyotype study experiments, enzyme extractions and determination of enzyme kinetics.

OTHER ACADEMIC AWARDS AND HONOURS

- Distinguished Graduate Student Award, 2012, Dept of Biological Sciences, Purdue University.
- Felicitated by the Vice Chancellor of Mumbai University, India for university merit ranking.
- Awarded scholarship from Mumbai University, India for standing 2nd at the University Level.
- Merit award for the College Honors Program at St. Xavier's College, Mumbai, India.

PROFESSIONAL SOCIETIES

- Chartered member of Society for Advancement of Biology Education Research (SABER).
- Member of American Physiological Society (APS).
- Member of Purdue International Biology Education Research Group (PIBERG).

SERVICE TO THE COMMUNITY

- Appointed as Brand Ambassador for Purdue University.
- Organized and disseminated a two-day workshop, *"Mini-grant: Investigating Students' Scientific Reasoning about Biological Experiments"* in Fall 2010 to bring together biology education and research faculty in the Mid-west region to participate in students' assessment design and evaluation for scientific reasoning about biological experiments.
- Graduate Research Assistant (2009) for evaluation of faculty response data for the research-based Diagnostic Question Clusters (DQCs).

REFERENCES

Nancy J. Pelaez (Advisor)
Department of Biological Sciences
Purdue University
West Lafayette, IN 47907-2054
Phone: 765-496-3261
Lab: 765-496-3092
Fax: 765-494-0876
Office: Lilly G-224
Email: npelaez@purdue.edu

David C. Eichinger (Chair)
Department of Curriculum and Instruction
Biology Education
Purdue University
West Lafayette, IN 47907-2056
Phone: 765-494-0711
Fax: 765-494-0239
Office: WTHR 221B
E-mail: deich@purdue.edu

Trevor R. Anderson
Department of Chemistry
Chemical Education
Purdue University
West Lafayette, IN 47907-2056
Phone: 765-494-5453
Office: WTHR 230B
E-mail: ander333@purdue.edu