Inspiration Brief 3

Enhancing Perceived Challenge/Skill and Achievement (DiSC 2005)

Debbie Denise Reese

NASA-sponsored Classroom of the Future

Center for Educational Technologies®
Wheeling Jesuit University

COTF/B3/Feb-2006

About the NASA-sponsored Classroom of the Future

The NASA-sponsored Classroom of the Future program is helping to bridge the gap between America's classrooms and the expertise of NASA scientists, who have advanced the frontiers of knowledge in virtually every field of science over the last 40 years. The program is administered by the Erma Ora Byrd Center for Educational Technologies® at Wheeling Jesuit University in Wheeling, WV

The Classroom of the Future™ serves as the National Aeronautics and Space Administration's premier research and development program for educational technologies. In this capacity the Classroom of the Future develops and conducts research on technology-based learning materials that challenge students to solve problems by using datasets and other information resources provided by the four core scientific missions underlying the work of NASA: Exploration Systems, Space Operations, Science, and Aeronautics.

About the Author

Report author Reese and the Inspiration Challenge staff are members of the Classroom of the Future research and evaluation team.

Dr. Debbie Denise Reese
Educational Researcher and Inspiration Project Lead

COTF 2006 Inspiration Challenge Staff

Lisa McFarland, Charlie Heiby (student), Katie Figaretti
With assistance from Greg Watson, Judy Martin

Project Consultants

Dr. James Schreiber
Alfred P. Sloan Working Family Center on Parents, Children, and Work: Dr. Barbara Schneider, Nora Burge, and Demetria Proutsos

Acknowledgments

Dr. Shelley Canright, NASA Education
Tammy Rowan, NASA Education
Dr. Charles A. Wood, Center for Educational Technologies
SPSS Statistical and Technical Support (David P. Nichols, Tom Leuthner, and Karen Gorny)

While this research project has benefited from the guidance provided by its consultants, advisors, program officers, and leadership, the author takes full responsibility for methodological decisions and interpretations derived thereof.

Table of Contents

Executive Summary

For NASA to succeed at building programs and technologies that inspire children, the agency must know what inspiration is, how to measure it, and more importantly, how to enhance it. To address this need, the NASA-sponsored Classroom of the Future (COTF) at the Center for Educational Technologies® at Wheeling Jesuit University in Wheeling, WV, launched an inspiration research agenda in 2005. The COTF defined the Systemic Model of Inspiration Growth, developed hypotheses of how to enhance inspiration through technology tools, built prototypes for two computer-mediated instructional technology tools, and conducted the baseline year of the inspiration lab research studies to inform tool refinement during subsequent cycles of design, testing, and revision.

Brief 3 reports the results of baseline testing of the DiSC (Discussion in a Scientific Context) inspiration tool conducted by COTF from September-December 2005 with more than 50 NASA Explorer Schools educators and 1,000 middle school students. A diverse demographic of students and their teachers from classrooms across the continental United States and Hawaii participated in the study. The brief also reports the effectiveness of Operation Montserrat, a NASA-approved instructional unit. Operation Montserrat is an e-Learning adventure because the four-week unit of classroom instruction culminates in a live simulation conducted via the Internet. During the two-hour simulation students work as scientists on crisis teams analyzing authentic data and responding to a hurricane/volcano disaster that actually occurred in 1996 on the island of Montserrat. The simulation is referred to as the e-Mission™. Inspiration 2005 was conducted with a fairly large sample of students, so COTF was able to identify trends in the data, even when they were very subtle. Results from the baseline study suggest:

1. Student perception of skills and challenges is higher during the e-Mission than at any other time during the four-week unit of classroom study. The literature identifies a state in which a person's skills and challenges are higher than his/her average as "flow." This effect was significant and modest.
2. Parents' level of education appears to have affected how DiSC prepared learners for the e-Mission. Students who reported their parents had completed high school or fewer years of education perceived higher levels of skills/challenges during the e-Mission when they had used DiSC. This effect was significant and modest.
3. Overall, the DiSC tool increased learners' perception of skills and challenges during the e-Mission. This effect was significant and weak.
4. Operation Montserrat increased student academic achievement an average of 1.5 points on a 16-item on a pre-/posttest. This was a significant and modest effect.
5. Learners with higher level of perceived skill and challenge during the e-Mission scored higher on a standards-based posttest drawn from national and state tests. This effect was significant and weak.
6. The COTF developed an argumentation self-efficacy scale for this study. Internal reliability for this scale was high (α_{pre}=.86 , α_{post}=.91).

Although these promising exploratory results support the COTF inspiration model and the research hypotheses, some are weak effects. COTF's Inspiration 2006 activities will include on-site facilitation of DiSC 2006 to increase control over study implementation fidelity. Enhanced implementation of the study instruments, instruction, and the DiSC tool itself should increase the effect of the DiSC tool upon learners' self-efficacy, mental models, and flow. COTF will also develop research instrumentation to support its claim of a stable measurement of flow during the e-Mission.

Background

In collaboration with Dr. Shelley Canright and Tammy Rowan of NASA's Technology and Products Office during 2004, the Classroom of the Future (COTF) at the Center for Educational Technologies® (CET) at Wheeling Jesuit University in Wheeling, WV, chartered a vision— a research agenda to define STEM-related inspiration. This included a definition of inspiration; hypotheses of how to enhance inspiration; design and development of prototypes for inspiration instructional technology tools based upon those hypotheses; and testing and refining those tools, the hypotheses, and the inspiration model through cycles of theory-based design and testing within inspiration labs. This report is a result of that vision as realized within COTF's 2005 cooperative agreement. It is the third in a series of COTF's Inspiration Briefs:

- Inspiration Brief 1 (Reese, Kim, Palak, Smith, & Howard, 2005) presented a literature review relevant to a construct of inspiration, a theoretical Model of Systemic Inspiration Growth, a set of initial inspiration hypotheses, and initial specifications for two inspiration lab studies.
- Inspiration Brief 2 (Reese & McFarland, 2006) elaborated the literature review specific to informal education and its two inspiration tools, overviewed the design and development of instruments to measure changes in inspiration, discussed the characteristics of COTF's two inspiration tools (the inspiration affective tool, RoboKids, and the inspiration social tool, DiSC—Discussion in a Scientific Context), presented the results of the inspiration lab baseline study that explored the effect of the RoboKids upon learners participating in an informal event, and discussed the implementation of the Inspiration Challenge baseline study of the DiSC tool.
- Inspiration Brief 3 (this report) contains the results of the Inspiration Challenge, a baseline exploratory study of the DiSC tool.

The Inspiration Construct

The COTF Model of Systemic Inspiration Growth (Reese, Kim, et al., 2005; Reese & McFarland, 2006) identified a set of five dimensions (identity, mental models, self-efficacy, imagination, and creativity) and flow that function together as a system to enhance inspiration. They are hypothesized to affect each other and an individual's propensity to make positive life choices. According to the model, the five parts work together to increase inspiration. Learners must grow along the five dimensions to increase their inspiration and make productive life choices that result in STEM literacy and pipeline achievement. When this happens, the result is an increase in flow, which COTF identified as a proxy for Inspiration. Inspiration Brief 1 (Reese, Kim, et al., 2005) contains a literature review for each of the dimensions and flow. As applied within learning environments, they can be summarized as:

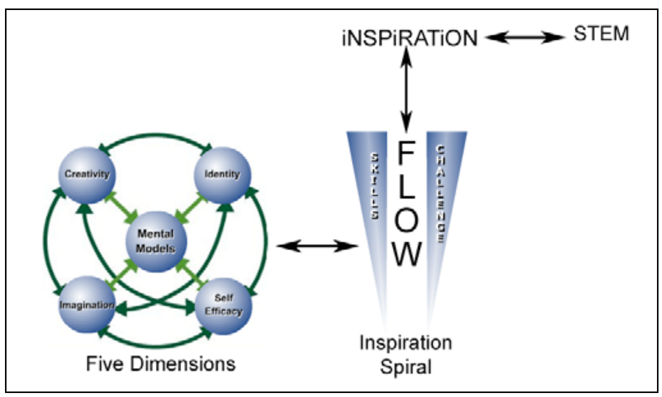


Figure 1. The Model of Systemic Inspiration Growth. From *Inspiration Brief 1: Defining Inspiration, the Inspiration Challenge, and the Informal Event (concept paper),* by D. D. Reese, B. Kim, D. Palak, J. Smith, and B. Howard, 2005, p. 10. Copyright 2005 by Center for Educational Technologies. Used with permission of the authors.

Dimension 1. Mental model.

Learners must construct mental models of science content and science enterprise (nature of science). A mental model is a person's understanding of a topic (Gilbert & Boulter, 2000). It is private, existing only in an individual's mind. But individuals can share what they know with other people when they talk about it, write about it, draw pictures about it, or make things based upon it. Mental models that agree with viable science knowledge are essential to science inspiration.

Dimension 2. Identity.

Each learner must construct self-image as someone who can do science individually and with others. Identification is "being recognized as a certain 'kind' of person in a given context" (Gee, 2001, p. 99).

Dimension 3. Imagination.

Each individual must move beyond the time and space of the present to invent and refine an identity as someone who can do science. Within classrooms or teams individuals must work together to invent themselves as a community that can do science. Individuals and the class must also invent solutions to science problems.

Dimension 4. Creativity.

Individuals must invent ideas and things that they never thought of before. They must do this by sharing their mental models of science with their classmates, team, or learning community and gaining the group's approval that the ideas are sound.

Dimension 5. Self-efficacy

As individuals and as learning communities, young people must come to believe that they CAN accomplish their science goals. Efficacy is a self-perception. It can also be an individual's or a group's perception of the group's collective ability to accomplish a task. This perception of group efficacy is known as *collective efficacy*. COTF presented an introduction to the self-efficacy literature in Brief 1 (Reese, Kim, et al., 2005). COTF discusses efficacy more extensively within Brief 2 (Reese & McFarland, 2006), where it provides the theoretical framework for the RoboKids Inspiration affective tool.

Flow.

Flow (Csikszentmihalyi, 1990, 1997; Csikszentmihalyi & Schneider, 2000) is an individual's immersion in experience characterized by intense focus, clarity, attention, and absorption directed toward accomplishment of the task at hand. During this period the individual's skill level must be equal to the challenge at hand, and both challenges and skills must be at optimal and relatively high levels (i.e., relative to the individual's mean level for the period of observation or comparison). Flow experiences are often accompanied by an altered perception of time and a connection with the activity that overrides a consciousness of self and ego. Flow is intrinsically rewarding, but re-occurrences of flow require a spiral of increasingly enhanced and relatively equal levels of skills and challenge.

DiSC 2005: The Inspiration Challenge

The DiSC 2005 Inspiration Challenge study used an experimental, pretest-postttest design to study the effects of the DiSC tool. The study measured changes in participants' mental models of the nature of science and scientific inquiry, science content knowledge (academic achievement), and argumentation. It also measured change in participants' argumentation, social, and academic self-efficacy. The overall goal of inspiration tools is to increase students' perceived flow. An individual's level of flow is operationalized as the individual's perceived level of skills and challenge and measured through the experience sampling method (ESM). COTF collected participants' perceptions of skills and challenges throughout the study using an ESM form. COTF hypothesized that DiSC would enhance learners' mental models of targeted content (academic achievement), argumentation, and practice of science; learners' argumentation self-efficacy; and flow. DiSC 2005 is an exploratory study designed to provide baseline data to be used in formative evaluation of DiSC, the development of design principles, tool revision, and refinement of research methodology. COTF developed or adapted five instruments for the study.

DiSC (Discussion in a Scientific Context).

Science reform standards emphasize that learners must engage in scientific discussions (argumentation) with their peers as they learn science content knowledge and skills because argumentation is authentic and engaging science practice. Scientific discussion is unfamiliar and challenging for both students and their teachers. DiSC is a web-based tool that introduces learners and their teachers to argumentation and guides learner teams as they practice argumentation. The theoretical framework for the DiSC tool is presented in Brief 1 (Reese, Kim, et al., 2005) and elaborated with development details within Brief 2 (Reese & McFarland, 2006).

Instructional Unit: e-Mission Operation Montserrat and the Live Simulation

The Inspiration Challenge used the Operation Montserrat (http://emissionhq.com/) as the targeted instructional unit. Operation Montserrat is a four-week unit of instruction from the Challenger Learning Center® at the Center for Educational Technologies. This e-Mission™ culminates in a two-hour live simulation conducted with participating classrooms via videoconferencing. During the simulation learners are placed into a disaster situation during which they must appraise the threats posed by a hurricane approaching the Caribbean island nation of Montserrat and an erupting volcano on the island. Students must plan an evacuation response that saves the inhabitants of Montserrat. Informal observations by Challenger staff support Center for Educational Technologies claims that the live simulation is a challenging experience that requires participating students to transfer skills and knowledge learned during Operation Montserrat toward a novel, authentic, real-time problem. DiSC 2005 used the experience sampling method to quantify students' perceptions of their level of skill, challenge, and other dimensions and qualities of experience during the e-Mission. The DiSC tool was designed to scaffold students' ability to work as a team to use evidence and reasons to make decisions. The e-Mission is a suitable transfer context for the argumentation skills and knowledge scaffolded by DiSC because the e-Mission requires students to work in teams to use evidence and reasons to make decisions.

Characteristics of the NASA Explorer Schools Pathfinder Initiative

Explorer Schools is a designated NASA pathfinder initiative designed to provide professional development within a three-year program that immerses underserved/underachieving schools in NASA science (National Aeronautics and Space Administration Office of Education, 2003). The program began in 2003 with 50 schools and adds a new cohort of 50 each year. School teams receive training and curriculum materials during summer workshops at NASA field centers. They are expected to disseminate training and materials throughout their building. The program provides onsite support at each school throughout the three school years. Major program goals concern ongoing professional development consistent with state and national standards, family involvement, technology integration, and student growth in (a) ability to apply STEM concepts, (b) knowledge about STEM-related careers, and (c) interest and participation in STEM.

Inspiration Brief 3 Scope and Purpose

The DiSC 2005 Inspiration Challenge study was designed to provide baseline data informing a research agenda addressing the DiSC tool inspiration hypotheses and studying the effectiveness of the NASA-approved product, e-Mission Operation Montserrat. The study targeted a population of learners from underserved, underachieving schools. Thus, it was important to begin to look at the relationship between the DiSC tool and an indicator of socioeconomic status that might affect student achievement within the inspiration model. In this case we isolated the variable of parents' level of education as a student-level indicator of human capital.

Hypothesis 1: DiSC Tool Inspiration

DiSC usage will increase student achievement along flow and its dimensions: mental models and self-efficacy.

Hypothesis 2: DiSC Tool Inspiration

There will be an interaction between DiSC usage and human capital in predicting flow during a transfer context.

Hypothesis 3: NASA-approved product e-Mission Operation Montserrat

The e-Mission Operation Montserrat unit will increase student achievement along flow and its dimensions: mental models and self-efficacy.

Although Brief 3 will briefly review the study mechanics, these were covered extensively within Brief 2 (Reese & McFarland, 2006). The primary purpose of Brief 3 is to report study findings and their implications for DiSC research and development during the COTF 2006 contract. A brief literature review will set the context for the findings. It will:

- Outline two methods used to compute flow and specify the relationship among students' perceived level of skill, challenge, and flow.
- Describe the Operation Montserrat e-Mission live simulation as a transfer context.
- Establish Parents' Level of Education as a measure of human capital.
- Describe the relevance of human capital given a targeted participant population of NASA 2003 cohort Explorer Schools.
- Provide an orientation to design-based research and summarize how this methodology has been used in the development of argumentation tools.

This report presents 2005 DiSC study results with respect to DiSC tool usage, the Operation Montserrat unit of study, and their effect on flow, mental models, and self-efficacy. Human capital (i.e., parents' level of education) is included within the analysis to investigate the interaction between context and the instructional interventions (Greeno, Collins, & Resnick, 1996). Brief 3 does not replicate the reporting of the inspiration literature review (see Reese, Kim, et al., 2005; Reese & McFarland, 2006) or tool development (see Reese & McFarland, 2006). Brief 3 does not compare the participating NASA Explorer Schools to other U.S. middle schools. Brief 3 reports pre-/postchanges across Inspiration dimensions measured by DiSC 2005 instruments, but it confines its analysis of results to those that concern flow. Brief 3 analyzes a subset of the data collected during DiSC 2005.

Literature Review

Computing Flow

Flow is typically calculated from individual participant's self-report using a nine-item Likert scale on two indicators (self-perception of skills and self-perception of challenges, see Figure 2). Flow composites are calculated in one of two ways. Both were designed to capture the "theoretically proposed and the empirically validated aspects of the relationship between challenges and skills" (Csikszentmihalyi & Schneider, 2000, pp. 261-262). Theoretically, flow is one of four experience states (Csikszentmihalyi, 1990, 1996; Csikszentmihalyi & Schneider, 2000). Flow occurs when both skills and challenges are high and relatively balanced. Apathy is defined as a state in which both skills and challenges are low and relatively balanced. Anxiety occurs when challenges are relatively high and skills are relatively low. Relaxation occurs when skills are relatively high and challenge is low.

When you were beeped. . .

	Low								High
How **challenging** was the activity?	1	2	3	4	5	6	7	8	9
How **skilled** were you at the activity?	1	2	3	4	5	6	7	8	9

Figure 2. The two items from the Experience Sampling Method (ESM) form used to indicate level of flow.

The structure of the ESM signal-level data file (Csikszentmihali & Larson, 1987).

Researchers are most familiar with data files structured at the level of the individual participant. Within participant-level datasets each participant forms a case, and all the data for each case runs across, forming a row. Each variable is collected across participants in a column, and each repeated measure of that variable forms an additional column. In contrast, ESM data is compiled at the signal or individual ESM response level. Flow researchers may call this the beep, signal-level, or response-level data file because participants respond by completing an ESM form each time they are signaled/beeped. Within an ESM data file each set of responses from one ESM form is a row and forms an individual case. Each study participant may be represented by as many rows in the dataset as there were ESM form collections. If there were 30 beeps during the study (signals to complete an ESM form), then each participant might have up to 30 rows of data. Repeated measures are handled differently within an ESM dataset. Each variable is listed only once; data collected at time 1, time 2, time 3 through time X for that variable will all be entered in the same column. Table 1 illustrates a sample response-level data structure for two participants who responded to four beeps on three variables. Of course, a researcher would never create an ESM data file with just two participants and only four signals.

Within ESM response-level datasets it is important to provide unique participant identifies (see "ID" in Table 1). It is also important to create a variable that records which beep 9ESm form administration) corresponds to the each case of responses (see "beep" in Table 1).

Table 1.
Sample Structure of the ESM Response-level Data File

ID	Beep	Skill	Challenge	Var03
1	1	9	5	6
1	2	5	2	3
1	3	3	6	6
1	4	6	8	5
2	1	5	5	5
2	2	1	4	6
2	3	5	5	1
2	4	3	6	7

Flow: Derived from z scores and dummy coding.

The four states of experience can be calculated using the ESM response-level data file (Csikszentmihalyi & Schneider, 2000). The first step is to calculate a mean value for each participant for skill and for challenge. COTF used SPSS version 14.0.1 software to split the response-level data file by participant ID numbers, then calculated *z* scores for each individual on both variables. This is called *z* scores at the person level. These *z* scores were used to calculate the four states of experience. When skills and challenge were both greater than 0 (greater than an individual's mean score across all responses for that variable within the study), flow would be calculated as = 1. Otherwise, flow = 0. The same system was used for each of the other three states of experience.

Flow: Derived from the geometric mean.

The geometric mean is the product of n variables raised to the (1/n) degree. Calculation of flow requires only two variables, skill and challenge. Thus, it is the product of two variables raised to the ½: $Flow = \sqrt{challenge * skill}$ or Flow $= (challenge * skill)^{1/2}$. This method results in one continuous measure of flow that "increases as challenge and skill increase" and "decreases as challenge and skill become more discrepant" (Csikszentmihalyi & Schneider, 2000, p. 261). Csikszentmihalyi and Schneider assigned a flow measure to each individual pair of responses.

Human Capital

James Coleman (1988) distinguished three types of capital that explain variance in individual achievement within education. Financial capital refers to wealth or income (Portes & MacLeod, 1996) and physical capital such as tools and other "productive equipment . . . that facilitate production" (Coleman, p. S100). Social capital refers to social relations within and external to a family (Short, 1997, p. 56). Within a school, classroom, or team, social capital refers to the norms of that community, the exchange of information, and the amount of trust. Human capital is the skills and abilities that equip a person to succeed. . . . Within the family, following common usage in economics, human capital is approximated by parents' education as an indicator of the potential for a child's cognitive development. Other measures of human capital include [an individual's] education and credentials that certify one's abilities. (p. 56)

Coleman was interested in social capital as a mechanism for increasing learners' human capital through enhanced networks establishing productive norms, expectations, information exchange, and trust.

Learners will enter the DiSC 2005 study with a fixed amount of human capital, as measured by their parents' combined level of education. It is usually assumed that a family's financial and human capital predict youth access to social capital (Short, 1997), but cultural and community characteristics within any one social system can provide extensive social capital even when financial and human resources are limited (Portes & MacLeod, 1996; Short, 1997).

DiSC and Human Capital

DiSC is a social tool that scaffolds learners through the argumentation procedure (see Figure 3). DiSC is designed to amplify students' ability to discriminate components of the argumentation procedure and apply them. Through the interface learners work in teams to recognize evidence and reasons that support or refute claims. The interface is a concrete tool designed to enable learners to work through the argumentation process within a team as they publicly share their private mental models by physically selecting information from a knowledge bank and moving it into a category as a reason or explanation supporting or refuting a claim. The tool provides role models and a rubric (see Reese & McFarland, 2006, Appendix E: The DiSC Tool Rubric) to reinforce learner engagement and teamwork. DiSC is designed to increase team and classroom social capital by establishing and reinforcing argumentation norms, exchange of information, norms, expectations, information exchange, and trust. Increased social capital should accompany increased human capital as DiSC prepares individual learners with argumentation skills and enhances Operation Montserrat learning accomplishment through social vetting or private mental models.

(a)

(b)

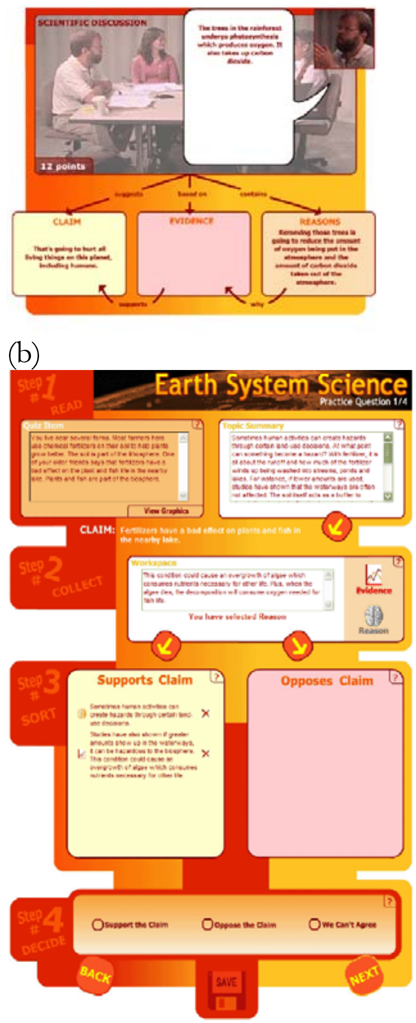


Figure 3. DiSC Tool Interfaces. Learners use the DiSC interface to drag and drop phrases into argumentation procedure categories. Interface module (a) helps learners to identify argumentation components and module (b) assists students as they discuss evidence and reasons that support or refute a claim.

Students who are initially well equipped for academic success probably have developed procedures for engaging with academic challenges. Learners who enter the DiSC environment with relatively high human capital, financial capital, and social capital may regard DiSC norms and procedures as an imposition/limitation because these students have already established a set of procedures within which they individually build academic successes. They may have to assimilate DiSC into already successful idiosyncrasies or accommodate DiSC by modifying existing academic practices. Because of the nature of the argumentation intervention, DiSC 2005 was designed to place students in situations (argumentation and the Operation Montserrat live simulation) that required teamwork with classmates. In contrast, students initially less endowed with financial, human, and social capital may not have developed strong methods for STEM study. Thus, COTF expects an interaction between existing human capital (parents' education) and DiSC usage. We would expect to observe this interaction when learners are placed in a situation that challenges them to transfer skills and knowledge from a unit of instruction toward a novel problem (for a review of transfer from the behaviorist/empiricist view, the cognitive/rationalist view, and the situative/pragmatist-sociohistoric view, see Greeno et al., 1996; for a discussion of an alternative approach to transfer via design experiments, see Lobato, 2003; for a brief introduction to transfer, see Reese, 2003, pp. 70-71; for a thorough treatment of the traditional definition of transfer, see Singley & Anderson, 1989). The Operation Montserrat e-Mission is a transfer task that challenges learners to work in teams as they apply the Earth system knowledge and skills toward emergency response to an authentic natural disaster. The learners are expected to abstract rules, principles, and schema from one learning context (i.e., classroom readings, discussions, and projects about volcanoes and hurricanes] and apply it toward another (i.e., the live simulation of the combined forces of the hurricane and the volcano on Montserrat). Students with existing high levels of capital might expect to use a personal arsenal of tried and true techniques to problem solve the e-Mission challenges. Students with low levels of existing human capital might not have developed a personal arsenal of problem-solving tools. DiSC 2005 will use parents' level of education as a measure of existing (exogenous) human capital and student perception of skill/challenge (flow) as a measure of human capital during the e-Mission experience. COTF expects that DiSC will enhance learners' level of flow during the e-Mission for students' with less human capital. During this four-week exposure to DiSC and argumentation, students with greater existing human capital may report less flow if they have used the tool with their classmates.

The effect of parents' human capital is limited by the quality of parent/child family networks (Coleman, 1988; Short, 1997). COTF did not control for family social capital, and this might reduce the effect of any observed interactions between human capital and DiSC usage.

Context, Flow, and Transfer

Unlike physical sciences that deal with closed systems (such as the movement of planetary bodies), in which predictions can be made with relatively few variables (Luke, 2004), the relevance of complex context in social sciences such as education motivates a perspective that considers exogenous influences. Although the unit of analysis might be the individual, systemic effects might be the effect of a context variable. Thus, this analysis included a statistical technique, mixed models, that allows researchers to control for context effects to confirm results from examining the effect of the interaction between tool usage and human capital on flow.

Human Capital and Explorer Schools

The NASA Explorer Schools program targets schools with underserved and underachieving populations of students. In 63 percent of the schools within the 2003 Explorer Schools cohort, a minority population comprised the majority of the school's students (Hernandez-Gantes, McGee, Reese, Kirby, & Martin, 2004). Schools were often located in areas that were economically depressed. COTF conducted and analyzed focus groups with the 2003 Explorer Schools during their first year in the program. Analysis suggested:

> The issue of poverty and the interplay with school culture was voiced by more than half of school teams (55 percent) who participated in focus groups. In some instances they spoke of

> the dire living conditions in the community: "Many of the children [in our school] come from homes with no running water or heat." In other cases they talked about the bleak prospects for students because of the lack of skills and incentives to perform well in school. For example, one teacher expects about a third of the middle school class to drop out "because of the lack of skills" and hopes that this can be overcome by getting students interested early in science, math, and technology. (p. 12)

Thus, the issue of the interaction between human capital and the DiSC tool should be a salient issue to investigate for the population of students represented by Explorer School educators who volunteered to participate in the Inspiration 2005 DiSC study.

Tool Design and Development through a Program of Research

Design-based research methods (Kelley, 2003) address a need for "approaches that speak directly to . . . the problems and issues of everyday practice" (Design-Based Research Collective, 2003, p. 5) and concern cycles of "design, enactment, analysis, and redesign" (p. 5) in which the design of learning environments and the development of theories are intertwined.

DiSC and GenScope.

DiSC derived from assessment and design-based research begun by Daniel T. Hickey and his colleagues (Hickey, Kindfield, Horwitz, & Christie, 2003) during a collaboration between the GenScope development team (originally at Bolt, Beranek, and Newman Labs, later relocated to the Concord Consortium) and an external team of assessment specialists at the Educational Testing Service. Following a multilevel assessment framework (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002), the assessment team developed Dragon Investigations— a set of proximal, formative assessment instruments (Hickey, Kindfield, et al., 2003). Although similar to proximal quizzes in the alignment between content and assessment items, Dragon Investigations included detailed answer explanations that teachers used as a follow-up in whole-class discussions with students. The answer explanations and discussions allowed teachers to implement the quizzes as formative assessments that provided achievement feedback to both students and their teachers (National Research Council, 2001; Smith & Ragan, 1993). The assessment team also used the multilevel assessment framework to develop instruments that served as proxies for high-stakes tests by randomly selecting genetics items from high stakes tests. Hickey (Hickey, Recesso, et al., 2003; Hickey, Zuiker, & McGee, 2004) continued to refine the design and implementation of the quizzes as well as the development of multilevel assessments. By 2004 Hickey (Hickey et al.) had elaborated the formative components of the quiz by incorporating the aspects of argumentation (Bell, 1997, 2004) and moving the student grouping from large group (teacher-moderated whole class) to small group (unmoderated discussion by about four students teamed for the instructional unit; see Hickey, Recesso, et al., 2003; Schafer, Kruger, & Hickey, n.d.; Taasoobshirazi & Hickey, 2005). COTF contracted with graduate students[1] at the University of Georgia trained by Hickey to develop three levels of assessment tools specific to Operation Montserrat: quizzes, a curriculum-oriented exam, and a test that served as a proxy for high stakes state and national instruments (see Appendix D). COTF later revised the quizzes and incorporated them into the DiSC tool. DiSC contains one set of practice quiz questions and topic summaries and three sets of curriculum-aligned quizzes.

[1] University of Georgia Learning and Performance Support Laboratory (LPSL) Operation Montserrat Assessment Team: Dionne I. Cross, Gita Taasoobshirazi, Kate Anderson, Steven J. Zuiker.

DiSC and SenseMaker.

Concerned with how "classroom debate and argumentation activities can help students integrate their scientific knowledge" (1998, p. 2), Philip LaVerne Bell explored how a "principled refinement process" (p. 2) might be used to refine a technology-based argumentation tool (SenseMaker) through cycles of "design, enactment, analysis, and redesign" (Design-Based Research Collective, 2003, p. 5). Bell's research was conducted within the framework of larger projects, such as Computer as Learning Partner (CLP), and products, such as Knowledge Integration Environment (KIE) and second generation Web-based Inquire Science Environment (WISE). These environments were designed to scaffold students' science knowledge integration through technology "learning partners" (Bell, 1998, p. 9) that were developed through design principles, while at the same time iterations of tool/curriculum design informed the theory that led to the design principles that would guide the design and development of those technology tools. By 2004 Bell had completed six cycles of design-based research. Each cycle informed one or more specific design principles, 11 over the six cycles. Bell and his colleagues used the design principles to refine the SenseMaker tool and its implementation as a learning environment within classrooms.

A Synthesis

In their review of educational research specific to cognition and learning, Greeno, Collins, and Resnick (1996) summarized and contrasted three psychological research perspectives:

- **Behaviorist.** Empiricism emphasizing "consistency of knowledge with experience" (p. 16) with the emphasis on efficiency in "conveying information and training skill" (p. 40).
- **Cognitive**. Rationalism emphasizing conceptual coherence and formal criteria of truth" (p. 16) with a focus on teaching as coaching (p. 40).
- **Situative**. Pragmatism emphasizing "that knowledge is constructed in practical activities of groups of people as they interact with each other and their material environments" (p. 16) with a focus on "teachers as mentors who represent communities of practice within the society" (p. 40):

> As such, [teachers] engage in the professional activities of creating and using disciplinary knowledge, exemplify valued practice of these communities, and guide students as they become increasingly competent practitioners. (p. 40)

Transfer, the application and elaboration of learning from one experience to others, is a major component of learning. Behaviorists track transfer as accomplished through gradients of similarity from prior knowledge toward a new situation that allows the individual to call up and apply previously associated responses. Within the cognitive perspective, learners abstract a schema (set of concepts and relationships) that is invariant across situations and apply it accordingly. Researchers in this tradition have found that learners often do not transfer learning from one education problem or situation to another unless that connection has been explicitly scaffolded. The authors suggest that learners must realize that the solution to any one problem is a general method that must be applied according to general features within relevant situations. They assert that programs that are successful at engaging learners in transfer "emphasize the social processes of explanation, formulation of problems and questions, and argumentation" (Greeno, Collins, & Resnick , p. 23). Within the situative perspective transfer entails that learners become "attuned to constraints and affordances through participation" (p. 23) while developing an identity as an engaged participant in the practice of the domain. For transfer to occur, learners must be able to recognize or acquire the constraints and/or affordances that are invariant from the original to the transfer situation.

The authors suggest that educational research should unite "individual, social, and environmental factors" (Greeno, Collins, & Resnick, p. 15) to provide "principles of a practical theory" that can be used to inform practice. Developed through design experiments, these principles will inform the formulation of curricula and assessments. Within the situative perspective, curricula provide the opportunity for participation in social practice that develops disciplinary discourse and representation (p. 36) through inquiry and discourse about concepts, claims, and questions. Activities should "include formulating and evaluating questions, problems, hypotheses, conjectures, and explanations, and proposing and evaluating evidence, examples, and arguments" (p. 31).

However, transfer and assessment have been problematic in situative, design experiments. "When knowing in a domain is considered as ability to participate in the socially organized distributed practices of thinking and inquiry in the domain, assessment needs to be focused on evaluation of those abilities" (Greeno, Collins, & Resnick, p. 39). The authors advise that achievement tests, criterion-referenced tests (those that compare learner performance to "explicit expectations," p. 38), and alternative assessments that "do not include sustained work on complex problems, communication, or collaboration with other people, or complex interaction with complex mechanical or other environmental technologies" (p. 38) are more aligned with a behaviorist perspective than a situative perspective. Situative assessments should "involve engagement with other people and with tools and artifacts that create natural, or 'authentic,' situations of activity" (p. 39).

Research Questions Restated

Results from the 2005 study would provide baseline insights that could be used to refine both the tool and implementation of study protocols. Specifically, COTF asked:

- Will the DiSC tool enhance inspiration along the mental models and self-efficacy dimensions and as operationalized by flow?
- Will e-Mission Operation Montserrat enhance academic achievement and flow?
- How does human capital interact with DiSC usage flow?

Methods

From September-December 2005 the NASA-sponsored Classroom of the Future (COTF) used the Experience Sampling Method (Reese, Kim, et al., 2005) to collect a total of 22,188 usable repeated self-reports of experience from 971[2] middle school students who were engaged in an Earth system science unit of study: e-Mission Operation Montserrat. Data collection was designed to quantify students' perceptions of their experiences with (a) COTF's newly developed DiSC (discussion in a scientific context) inspiration instructional tool and (b) COTF's Operation Montserrat live simulation e-learning adventure. In addition, COTF collected academic achievement data and measures of students' mental models of science, argumentation, and their self-efficacy. COTF had hypothesized that both DiSC and the e-Mission would enhance students' inspiration as projected within the COTF Model of Systemic Inspiration Growth (see Figure 1).

Participants

Fifty NASA Explorer Schools middle school classes were randomly assigned to treatment and control conditions at the classroom level. Students from the 45 classrooms (36 teachers) actually began the 2005 DiSC Inspiration Challenge study in September 2005 (see Appendix A, school-level

[2] A total of 997 usable student records were collected across all study instruments. Some students submitted no valid ESM data, reducing the number of students with ESM records to 971.

data derives from National Council of Education Statistics, NCES, http://nces.ed.gov/ccd/schoolsearch, 2002-2003 reports unless otherwise noted in the appendix). Thirty-nine classrooms (31 teachers) successfully completed components of the study. Each teacher who submitted the entire set of study instruments (test, exams, four surveys, and a majority of ESMs) received a $200 stipend per participating classroom. Data scanning and cleaning resulted in 997 student cases with usable data. Of these, 30 classrooms (26 teachers) completed the study with implementation fidelity in use of the DiSC tool. DiSC tool usage analyses use data from the 716 students enrolled in those classes (the tool usage dataset).

Students within the tool usage dataset attended schools located in Arizona $n_{student}$ (n=75), California ($n_{student}$=121), Florida ($n_{student}$=30), Georgia ($n_{student}$=33), Hawaii ($n_{student}$=29), Iowa ($n_{student}$=35), Kentucky ($n_{student}$=60), Massachusetts ($n_{student}$=34), Mississippi ($n_{student}$=53), New Mexico ($n_{student}$=20), Nevada ($n_{student}$=62), New York ($n_{student}$=21), South Carolina ($n_{student}$=66), Texas ($n_{student}$=45), and Vermont ($n_{student}$=32). School enrollments ranged from 219 to 1,262, and schools were situated in diverse locales: isolated/rural ($n_{student}$=104), rural ($n_{student}$=193), suburban ($n_{student}$=194), urban ($n_{student}$=192), and urban-inner city ($n_{student}$=33). All of the students attended public schools, with 30 attending a public magnet school.

Socioeconomic status (SES) and human capital.

The majority of students in the tool usage dataset attended schools that reported no migrant workers ($n_{student}$=425). The remaining students attended schools with 1 to 65 migrant families. COTF determined socioeconomic status (SES) as a ratio of the sum of free and reduced price lunches to total student body. SES ratios at the DiSC-usage school ranged from 20 percent ($n_{student}$=63) to 100 percent ($n_{student}$=20). About 67 percent of these students attended schools at which 50 percent or more of the student body qualified for reduced or free lunch.

Parents' level of education is typically used as a measure of students' human capital (see discussion of human capital within the Literature Review section of this brief). According to student report of parents' education levels with presurvey 1 (see Table 2, Appendix B, and Figure 1), 23 percent of students come from families in which parents' level of education is high school and 23 percent from families in which parents' level of education is college. A composite of father's/mother's education (Parents' Level of Education), $\sqrt{(education_{mother})(education_{father})}$, ranges from 7.4 percent reporting less than high school to 1.9 percent reporting the Ph.D. or M.D. level.

Table 2.
Number of Students By Self-reported Level of Parent's Education

Parent Education[1]	Father	Mother	Composite[2,3,4]
1 (< high school)	61	70	40
1.41	—	—	32
1.73	—	—	15
2 (high school only)	226	185	127
2.24	—	—	1
2.45	—	—	91
2.83	—	—	20
3 (college)	195	220	131
3.16	—	—	8
3.46	—	—	32
3.87	—	—	10
4 (master's)	61	70	32
4.47	—	—	12
5 (Ph.D. or M.D.)	23	33	11
Missing data	150	138	154

Notes: [1] Value labels pertain to only mother's or father's education. Numbers apply to all three columns. [2]Composite Parents' Education Level is the geometric mean of a student's mother's and father's education level $\sqrt{(education_{mother})(education_{father})}$. [3]Scale translations for composite Parents' Education Level (pairs are not ordered): 1=(< high school ,< high school), 1.41=(< high school, high school), 1.73=(< high school, college), 2=(high school, high school) or (< high school, master's), 2.24=(< high school, Ph.D./M.D.), 2.45=(high school, college), 2.83=(high school, master's), 3=(college, college), 3.16=(high school, Ph.D./M.D.), 3.46=(college, master's), 3.87=(college, Ph.D./M.D.), 4=(master's, master's), 4.47=master's, Ph.D./M.D.), 5=(Ph.D./M.D., Ph.D./M.D.). [4]Composite calculations drop cases with any missing data.

Gender and grade level.

Within this tool usage dataset, 332 students were female and 396 were male, and 88 did not report their gender. All participating students attended middle school: grade 6 ($n_{student}$=282), grade 7 ($n_{student}$=139), grade 7 and 8 combined ($n_{student}$=49), and grade 8 ($n_{student}$=246). However, participating schools housed eight different grade-level structures: pre-kindergarten–grade 8 ($n_{student}$=17), kindergarten–grade 6 ($n_{student}$=27), kindergarten–grade 12 ($n_{student}$=20), grades 4-8 ($n_{student}$=53), grades 6-8 ($n_{student}$=447), grades 7-8 ($n_{student}$=90), grades 6-12 ($n_{student}$=35), and grades 7-12 ($n_{student}$=27).

Study Design

A detailed description of the study design is available in Inspiration Brief 2 (Reese & McFarland, 2006), but will be summarized here (see also Appendix G: DiSC 2006 Experimental Design). The Inspiration Challenge design was a pretest/posttest randomized experiment. Fifty teachers were randomly drawn from 70 self-selected, NASA Explorer Schools from its 2003 cohort (the program's first year). COTF staff randomly assigned classes to treatment (use of the DiSC tool to scaffold argumentation) and control (use of a placebo tool that had the same look and feel as the DiSC tool but did not scaffold argumentation) conditions. Each teacher's class was assigned a class ID number. Each teacher was provided with a set of student ID numbers. Each teacher assigned student numbers and logged them on an ID sheet. All student data was coded by students using student ID numbers. No student names were collected on data. This study was approved by the Wheeling Jesuit

University Institutional Review Board, and COTF retained data for only students who had submitted an informed consent form signed by a parent or guardian. Student consent forms were keyed to student ID numbers by an ID sheet completed by each participating teacher research partner. Teachers submitted ID sheets to the DiSC 2005 staff, and staff destroyed the sheets.

DiSC 2005 held everything constant for both conditions, except for the version of the tool used by students (DiSC or placebo). All classes were expected to follow the same schedule (see Figure 4) as they progressed through the Operation Montserrat unit of instruction:

- September 26-30: Baseline administration of the ESM form, administration of survey 1, survey 2, and the curriculum-oriented exam, and tool (DiSC or placebo) training.
- October 3-7: Operation Montserrat lessons 1-3, ESMs 6-11, Quiz 1, and tool session 1.
- October 10-14: Operation Montserrat lessons 4-6, ESMs 12-17, Quiz 2, and tool session 2.
- October 17-21: Operation Montserrat lessons 7-9, ESMs 18-23, Quiz 3, and tool session 3.
- October 24-28: Operation Montserrat lessons 10-12, ESMs 24-29.
- October 31-November 11: Conduct the Operation Montserrat live simulation e-Mission, complete ESM #30 (these both occur on one day); then complete survey 1, survey 2, curriculum-oriented exam, and the standards-oriented posttest over the course of the following four days.

Four COTF staff members provided daily to weekly implementation guidance to all participating teachers via telephone and e-mail. Challenger Learning Center staff were available to guide teachers in execution of the Operation Montserrat curriculum and technical setup for the live simulation via the Internet. The study was also supported by a Blackboard web site that contained all study instruments and implementation directions.

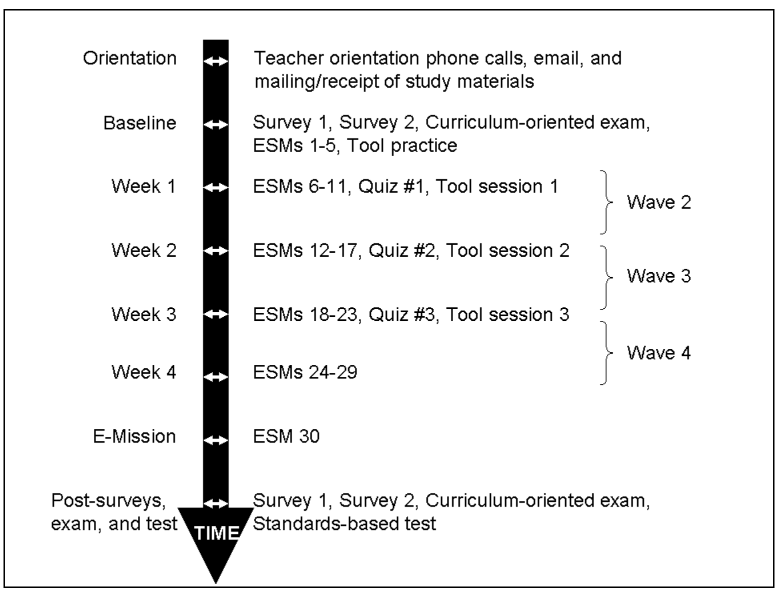


Figure 4. The Scheduled and Calculated Administrations of DiSC 2005 Instruments. DiSC 2005 was originally organized by week (see left-hand side of arrow). Because of teachers' idiosyncratic implementation schedules, the research team used shipment, processing, and facilitator records as well as consultation with individual teachers to group each classroom's ESM instruments into baseline, waves 2-4, and e-Mission (see waves on right-hand side of arrow).

Quizzes and DiSC tool.

The quiz/tool sessions structure derived from Hickey's multilevel assessment framework (Hickey et al., 2004; Taasoobshirazi & Hickey, 2005). DiSC 2005 scheduled for students to complete a set of lessons (e.g., week 1 lessons 1, 2, and 3) and take a quiz on that content (e.g., quiz 1). This much is typical practice within many classrooms. Within DiSC 2005, as within the Hickey framework, students break into teams after completing the quiz and discuss the answers (using their argumentation/scientific discussion skills). To this COTF added the COTF inspiration social DiSC tool as a technology partner (Bell, 1998) to scaffold discussion. Teams met at a computer station and interacted with DiSC or the placebo tool to discuss the quiz responses. Both DiSC and placebo presented the four quiz questions, one at a time. Each question was accompanied by a topic summary that provided information from the week's lessons that the team could use to develop an answer to that quiz question. Then the placebo tool asked each team member to record whether they agreed or disagreed with the claim (quiz question). The DiSC tool scaffolded argumentation by making the argumentation process visible:

- Each team selected components of the answer explanation and moved them into a workspace.
- The team modified any workspace text as desired and identified it as either evidence or reason.
- The team moved the workspace content and label into a category list as supporting or opposing the claim.
- The team logged whether, as a group, it supported the claim, opposed the claim, or could not as a team come to an agreement.

During the baseline week (see Figure 4), both the treatment (DiSC tool) and control (placebo) classes completed a practice session with their tool. In addition to practicing the quiz question interface, the DiSC group received training in identification of (a) proper argumentation practices (i.e., engagement; turn-taking; and science talk components of claims, evidence, and reasons) and (b) practiced in identifying claims, evidence, and reasons.

ESM administration.

Directions for administering each ESM (see Appendix C) were distributed to teachers in sealed enveloped labeled by the ESM number. Teachers were instructed to open either one or two envelopes each class period and set a timer (provided by CET) for the number of minutes listed within the instructions. Students were to complete the ESM when the timer rang. ESM directions included verbal instructions to be spoken by the teacher during ESM administration. Teachers were expected to administered the ESM form 30 times.

ESM 30 was the critical sampling of student experience for this study because it measured a consistent experience context for every student in the study: their experience at e-Mission time = 55 minutes. Timing of this ESM was controlled by the flight directors at the Challenger Learning Center who were conducting the e-Mission. Again, administration of ESM 30 occurred 55 minutes into the Operation Montserrat e-Mission for all students across all participating classes. Although each e-Mission is idiosyncratic because each scenario responds to decisions made by each participating class during the simulation, e-Missions are a scripted experience: same situation, same actor roles for student participants, and identical hurricane/volcano events over time. By time = 55 each e-Mission is mid-scenario and well under way. The only differences in implementation are students' emergency response team decisions. These decisions haven't begun to affect the e-Mission system by time = 55.

Random selection of the ESM timings.

ESM timings for ESMs 1-29 were randomly selected. Timings were randomly drawn from hat—with replacement. Possible timings for ESMs 1-29 were 5 minutes–45 minutes (allowing for a 50-minute class with 5 minutes to start the class and a few minutes to dismiss). Four days were scheduled for two ESM collections (week 1, day 2; week 2, day 3; week 3, day 4; week 4, day 5). The constraint for days with double sampling was that the two times had to fit within 45 minutes. For example, the final ESM sampling day sampled at 9 minutes into the class and at 32 minutes into the class, leaving enough time to complete the second ESM and dismiss the class. The e-Mission time = 55 was selected because it is about halfway through the mission.

The Inspiration Challenge Context

DiSC 2005 used the context of a competition to motivate teachers and students. All classrooms that completed DiSC 2005 were eligible to complete in the Inspiration Challenge for the Most Inspired Classroom. Entrance consisted of student-selected evidence of their class's inspiration during the Operation Montserrat unit of study along the five dimensions of the Model of Systemic Inspiration Growth (see Figure 1). Up to three treatment and three control class prizes could be awarded. An internal panel of CET judges who had not been involved with the DiSC 2005 study evaluated the students' artifacts. Each winning class received a plaque and $1,000 to be used by the school for the purchase of STEM classroom technology. COTF DiSC 2005 awarded two prizes, one to a treatment and one to a control classroom.

Instruments and Measures

DiSC 2005 used 30 administrations of its ESM form, two pre-/postsurveys, a pre-/postexam, and one posttest (see Table 3). The survey and the ESM instruments were developed or adapted from four sources (see Appendix E):

- The seven-year Sloan Study of Youth and Social Development conducted by the Alfred P. Sloan Working Family Center (http://www.sloanworkingfamilies.org/) at the University of Chicago and the National Center for Research (http://www.norc.uchicago.edu/).
- Norm G. Lederman's (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002) research and Views of Nature of Science questionnaire.
- Deanna Kuhn's (1993) research about the nature of science.
- Albert Bandura's academic and social self-efficacy scales and item-writing guidelines (Bandura, 2004).

This report uses only a subset of the DiSC 2005 instruments and data.

Assumptions.

1. The ESM captures self-perception of "a person's consciousness" (Csikszentmihalyi & Schneider, 2001, p. 122) at the experience level.
2. ESM responses are typically collected over a participant's waking hours over infinitely varied contexts. Experiences provided by diverse contexts are often very different, leading to volatile response patterns that are unstable across observations. ESM social scientists, interested in participant responses patterns across those diverse contexts, aggregate five to eight ESM responses over a period (what COTF terms a "wave") to ensure response stability. DiSC 2005 measures student response in a context that is relatively consistent: one school class throughout

the Operation Montserrat unit of instruction. Thus, DiSC application of ESM—especially for ESM 30— is similar to a repeated measurement via a standard instrument (e.g., a survey or test).

3. The measures of skill and challenge collected at e-Mission are "snapshot" (Csikszentmihalyi & Schneider, 2001, p. 122) indicators of each student's perceived level of skill and challenge midway into the e-Mission, at time = 55 minutes. Across all students this one-time measure of experience is a stable measure of mid-mission experience.
4. Composites derived from ESM 30 skill and challenge indicators can be used in either of the two flow calculations (see section "Computing Flow", this report).
5. Composites derived from ESM 30 skill and challenge indicators can be used in either of the two flow calculations (see section "Computing Flow", this report).
6. Flow and state of experience variables derived from DiSC 2005 ESM 30 can be used as either dependent variables or predictors of postmission achievement.
7. Other quality of experience indicators derived from DiSC 2005 ESM 30 are also indicators of experience at e-Mission time = 55 minutes.
8. Baseline DiSC data across all participating classes was collected while students were taking the pretests and surveys or practicing with the tool (DiSC or placebo). Thus, baseline ESM data should be relatively stable and require aggregation of fewer ESM responses.
9. Although ESM waves 2, 3, and 4 were also collected during the Operation Montserrat unit of instruction, there was more variability in teacher and student activity as teachers followed idiosyncratic implementation schedules and unit activities. Response stability issues support aggregation of a minimum of five or more ESM collections during DiSC waves 2, 3, and 4.

Although ESM 30 is assumed to capture an average indication of student experience at e-Mission time = 55, the results section will report using the terms "skill" and "challenge" rather than flow or the other three states of experience (apathy, anxiety, relaxation). Following from the assumptions within this section, the conclusion and discussion sections will transition from indicator terms to the four experience dimensions.

ESM organization into waves.

Although COTF had planned a defined schedule of events (instruction and instrument administration, see Figure 4), each teacher's implementation schedule was idiosyncratic because of district scheduling and factors such as weather conditions. DiSC staff used mailing records, data processing records, facilitator records, and teacher input to organize each teachers' set of ESM forms into baseline (testing and tool practice, BEFORE instruction began), wave 2, wave 3, wave 4, and e-Mission (see Figure 4). During analyses ESM data except e-Mission (ESM 30 is one observation) are filtered for number of observations: baseline > 2 and waves 2-4 > 4.

Parents' level of education.

Parents' level of education is a student-level measure derived from student report collected within the prestudy survey at the start of the DiSC study (see Table 2). Students indicated both their mother's level of education and father's level of education (1 = less than high school, 2 = high school only, 3 = college, 4 = master's , 5 = Ph.D. or M.D.). The derived variable is the geometric mean of these two scores; that is, the square root of (mother's education X father's education) calculated for each individual student.

Socioeconomic status.

Socioeconomic status (SES) is a school-level variable derived from National Center for Education Statistics (NCES) in 2002-2003 (the year in which each school entered the three-year NASA Explorer Schools program). SES is calculated for each school as the proportion of free and reduced price lunches within a school's student body. Instances in which NCES statistics were unavailable or replaced by information from alternative sources are noted within the tables in Appendix A.

Perceived challenge/skill.

Perceived challenge/skill is calculated in two ways (see Computing Flow section of this report).

1. *z* scores and dummy coding, specifically as[3]:
   - High challenge/High skill = (+ challenge/ +skill).
   - High challenge/Low skill = (+challenge/- skill).
   - Low challenge/High skill = (-challenge/+ skill).
   - Low challenge/Low skill = (-challenge/- skill).
2. An individual's geometric mean of (a) the aggregated baseline, wave 2, wave 3, and wave 4 skill and challenge indicators and (b) the e-Mission skill and challenge indicators. Within this report this indicator is labeled challenge/skill.

Academic achievement.

Academic achievement items were developed using Daniel T. Hickey's (2004) multilevel assessment framework. A team from his lab at the University of Georgia developed the first iteration of the quizzes, exam (Operation Montserrat curriculum-oriented exam), and test (standards-based test). Appendix D contains excerpts from the teams' final report on development of the assessments.

Operation Montserrat curriculum-oriented exam.

Within the multilevel assessment framework (Hickey et al., 2004) this 16-item assessment (see Appendix D) is categorized as a proximal because it, like many formal classroom assessments, was designed to "match the facts and skills targeted by the curriculum" (p. 6) for administration after an entire curriculum has been completed. The exam was administered during the baseline week (pretest) and after completion of the e-Mission (posttest). Internal reliability statistics are reported in Table 3. Appendix F contains difficulty and discrimination indexes for this exam. Table 3 lists internal reliability statistics. Exam items were taken from a pool of items written by professional test item writers that comprise two parallel forms of a 40-item Operation Montserrat exam (Howard, 2004). Pre- and posttests were completed by 508 students from classrooms in four states (West Virginia, New York, Ohio, Pennsylvania) who had completed Operation Montserrat instruction and the e-Mission (Reese, Smith, & McFarland, 2005). Some of those students had been targeted as living in impoverished counties (*n*=152), with Operation Montserrat e-Missions funded by a grant from American Electric Power of West Virginia and Ohio (AEP). Many of the participating teachers had completed a one-day Operation Montserrat pre-implementation training workshop, and some had taught the unit and conducted an Operation Montserrat e-Mission once before. Results from that report provide some comparison for the DiSC 2005 mean scores, given as a percentage of the 40 points possible on that test: $\overline{X}_{\text{pretest (AEP county)}}=33\%$, $\overline{X}_{\text{pretest (non-AEP county)}}=32\%$; $\overline{X}_{\text{posttest (AEP county)}}=36\%$, $\overline{X}_{\text{posttest (non-AEP county)}}=37\%$; $\overline{X}_{\text{gain (AEP county)}}=3\%$, $\overline{X}_{\text{gain (non-AEP county)}}=5\%$.

[3] '+' = greater than 0, '-' = ≤ 0

Standards-based test.

Within the multilevel assessment framework (Hickey et al., 2004) this 24-item assessment is categorized as a distal because it uses "quasi-randomly selected multiple choice items selected from a larger pool of items that reflect the targeted standard(s)" (p. 6). The test items were randomly selected from a pool of standardized test items related to Earth system science and aligned with science content standards. The test was administered after completion of the e-Mission (posttest only). Internal reliability statistics are reported in Table 3. Appendix F contains Difficulty and Discrimination Indexes for this test; Table 3 lists internal reliability statistics.

Table 3.
Internal Reliability Statistics for Curriculum-oriented Pre-/Postexam and Standards-based Test.[1]

	$\text{Curriculum-oriented}_{pre}$	$\text{Curriculum-oriented}_{post}$	Standards-based
K-R 20	.58	.71	.79
Spearman-Brown	.62	.72	.78
Guttman Split-half Coefficient	.62	.72	.77
$n_{subjects}$	761	698	610
N_{items}	16	16	24

[1] $N_{subjects}$=997

Academic self-efficacy.

DiSC 2005 used Bandura's (2004) six-item Likert academic self-efficacy scale (e.g., How well can you learn general mathematics?) rated from 1 (*can't do at all*) to 5 (*can do very well*). Cronbach alpha levels for all participants are α_{pre}= .68, $N_{participants}$=790 and α_{post}= .73, $N_{participants}$=729. Cronbach alpha levels for the subset of participants within the DiSC usage analyses are α_{pre}= .70, $n_{participants}$=557 and α_{post}= .75, $n_{participants}$=534. Participant score for this scale was the sum of the six items for a maximum score of 30.

Social self-efficacy.

DiSC 2005 used Bandura's (2004) seven-item Likert social self-efficacy scale (e.g., How well can you make and keep friends?) rated from 1 (*can't do at all*) to 5 (*can do very well*). Cronbach alpha levels for all participants are α_{pre}= .70, $N_{participants}$=725 and α_{post}= .71, $N_{participants}$=710. Cronbach alpha levels for the subset of participants within the DiSC usage analyses are α_{pre}= .70, $n_{participants}$=505 and α_{post}= .72, $n_{participants}$=521. Participant score for this scale was the sum of the seven items, for a maximum score of 35.

Argumentation self-efficacy.

DiSC 2005 used Bandura's (2004) guidelines to construct a scale to measure argumentation self-efficacy. The scale consists of nine Likert-type self-efficacy items (see Table 4), rated on a scale from 0 (*can't do at all*) to 4 (*can do very well*). Cronbach alpha levels for all participants are α_{pre}= .86, $N_{participants}$=761 and α_{post}= .91, $N_{participants}$=694. Cronbach alpha levels for the subset of participants within the DiSC usage analyses are α_{pre}= .86, $n_{participants}$=533 and α_{post}= .91, $n_{participants}$=508. Participant score for this scale was the sum of the nine items, for a maximum score of 36.

Table 4.
The DiSC 2005 Argumentation Self-efficacy Items.

1. I can tell others when my science ideas are correct.
2. I can tell others what evidence supports my science ideas.
3. I can tell others why that evidence supports my science ideas.
4. I can tell others when evidence does not support their science ideas.
5. I can ask others what evidence supports their science ideas.
6. I can ask others for reasons why their evidence supports their science ideas.
7. I can listen to ideas that don't agree with my ideas.
8. I can change my idea if other people have better evidence for their ideas.
9. I can tell when other people's evidence is better than mine.

Job knowledge: Operation Montserrat aligned occupations.
Job knowledge specific to Operation Montserrat-aligned occupation (e.g., volcanologist, meteorologist, geologist, emergency response teams) was measured by 18 items. Participants used a five item Likert scale (1=*strongly disagree*, 2=*disagree*, 3=*not sure*, 4=*agree*, 5=*strongly agree*) to answer statements such as "A volcanologist works only when volcanoes are active." Some items were written in reverse to obviate response set. Choices of (*agree* or *strongly agree*) were scored 1 point for correct statements. Choices of (*disagree* or *strongly disagree*) were scored 1 point for incorrect statements. One item was removed from scoring because its wording was unclear. Job knowledge score was a participant's sum correct across the remaining items, with a maximum of 17.

Nature of science.
Fifteen nature of science items (see Table 5) were adapted from the literature (Kuhn, 1993; Lederman et al., 2002). COTF synthesized the argumentation literature to develop two additional nature of science items featuring scientific discussion. Items were written to measure the degree to which the participant identified science as a dynamic and evolving evidence and theory-based practice that is a joint endeavor by the community of scientists. Items written in reverse were translated during scoring. A participant's nature of science score is the total sum score over the Likert 17 items: strongly disagree 0 = *disagree*, 2 = *not sure*, 3 = *agree, 4* = *strongly agree* . Maximum possible score is 68. Cronbach alpha levels for all participants are α_{pre}= .70, $N_{participants}$=720 and α_{post}= .77, $N_{participants}$=689. Cronbach alpha levels for the subset of participants within the DiSC usage analyses are α_{pre}= .73, $n_{participants}$=500 and α_{post}= .76, $n_{participants}$=494.

Approach and Software
The DiSC 2005 analysis is an exploratory study to identify trends, relationships, and effects that can be used (a) as a baseline with which to compare future iterations within this program of design research and (b) to identify tool and implementation areas to highlight or that require revision. SPSS 14.0.1 software was used to conduct all analyses. The direction of the analyses is informed by theory, but analyses are informed by the characteristics of the data itself. As such, many results are descriptive rather than predictive.

Table 5.
Nature of Science Items.

1. Scientists' beliefs and values do not influence their research. (rev)
2. After scientists have developed a theory (for example, atomic theory, evolution theory), the theory never changes. (rev)
3. Scientists perform investigations through experiments and observations.
4. Scientists do not use creativity and imagination when solving scientific problems. (rev)
5. Scientific knowledge is based on a natural event (for example, earthquakes, hurricanes), evidence, data, and observation.
6. All scientific practice is based in experiments. (rev)
7. Scientists' claims are influenced by the scientific and cultural environment of the time.
8. Scientists read and talk only about things that agree with their ideas. (rev)
9. Scientists are curious about the world in which we live.
10. Scientists like to listen to people whose ideas are different from theirs.
11. Scientists' ideas grow strong when other scientists question their reasons.
12. Scientists report unexpected results as well as expected ones.
13. Scientists report only results that agree with their ideas. (rev)
14. Discussion is central to science practice.
15. Scientists do not use discussion when they want to make other people agree with their ideas. (rev)
16. Scientists are unwilling to change their ideas when evidence shows that the ideas are poor. (rev)
17. Scientists' new ideas are not affected by where they live, whom they talk to, and what they already know. (rev)

Results

The results are presented in three major subsections. The first subsection summarizes results for the three academic achievement instruments: pre-/post-Operation Montserrat curriculum-oriented exam and the standards-based posttest. The second subsection summarizes results for the self-efficacy (academic, social, and argumentation), the Operation Montserrat-specific job knowledge and nature of science scales. Flow and the other three dimensions of experiences (apathy, relaxation, and anxiety) are operationalized as perceived levels of skill and challenge. The third subsection reports analyses of students' perceptions of challenge/skill levels during the Operation Montserrat unit of study at baseline, waves 2, 3, and 4, and at e-Mission (time = 55 minutes) as well as the interaction between tool usage (DiSC or placebo tool) and perceived skill/challenge. Human capital indicators are added to provide a more detailed picture of how these Explorer Schools students, targeted as low SES and achievement, interacted with DiSC. Because students are nested within classrooms, classrooms within teachers, and teachers within schools, a mixed model analysis is briefly summarized as a control for the effects of nesting and SES context. This section also looks at perceived challenge/skill as a predictor of academic achievement.

Although the DiSC tool was the focus of the Inspiration Challenge 2005 study, some participating teachers did not follow DiSC usage protocols. Facilitator communication monitoring teachers' implementation and our database management/reporting system identified illegitimate teacher and student implementation (see Table 6). Some examples are:

- The tool is designed to scaffold small group argumentation. Two teachers implemented the tool with their students as a whole class.

- Teachers in one school shared a mobile computer lab, and one class of students logged on to both versions of the tool.
- Some teachers did not implement all three components of the tool practice session with students. Tool practice was a baseline week activity and included a video section plus interaction with the tool.
- Some teachers did not conduct all four tool sessions.
- One teacher completed all study components except the tool.
- In some cases individual students or teams of students were removed from condition assignments because of their implementation of the study. For example, one team of students used both iterations of the tool because they logged into the incorrect tool one time.

Within this report, analyses of the total population of students include students of any teacher who completed instruments and the unit of instruction through the e-Mission (labeled “all participants”). Analyses pertaining to the DiSC tool use only students from classrooms with tool implementation fidelity. Their dataset is the “tool usage” data subset. The control condition is referred to as using the “placebo tool” or as the “placebo” condition. The placebo tool did not scaffold argumentation. The treatment tool did scaffold argumentation. The treatment condition is referred to as using the “DiSC tool” or as the “DiSC” condition. The third subsection uses only the tool usage data subset.

Table 6.
Tool Usage Implementation Characteristics and Specification of the Tool Usage Dataset for the DiSC 2005 Inspiration Challenge Study.

	$N_{students}$	$N_{classes}$	Tool Usage Dataset
Control			
Placebo tool	331	13	Yes
Tool as a class	30	1	—
Tool as class, no tool session 3	28	1	—
Never did tool	21	1	—
Used both versions of tool	27	1	—
Treatment			
DiSC tool	385	17	Yes
Video training but no tool practice	81	3	—
No video training and no tool practice	57	2	—
No condition assignment	37	—	—
Total	997	39	

Achievement

Operation Montserrat curriculum-oriented exam.

Means and standard deviation statistics for DiSC 2005 pre/post measures are listed in Table 7. All DiSC 2005 Operation Montserrat curriculum-oriented exam pretest scores are in line with expectations, based upon previous iterations of testing achievement using the larger pool of test items with students from AEP targeted and non-AEP targeted schools (see Table 8). Pretest DiSC percentages (DiSC = 33 percent, 34 percent) are very similar to pretest AEP/non-AEP percentages (32 percent, 33 percent). Gain percentages for the DiSC implementation are greater than gain within the AEP/non-AEP report. Many of the teachers who participated in the AEP/non-AEP Challenge had completed an Operation Montserrat training workshop and some had previously implemented

Table 7.
Means[1] and Standard Deviations for Measures Derived from Exams (Pre-/Post-Operation Montserrat Curriculum-oriented [OM] Exam), Tests and Surveys (Pre/Post) Across All Participants and for Control (Placebo) and Treatment (DiSC) Conditions of Tool Assignment.

	Pretest									Posttest								
	All Participants[9]			Control			Treatment			All Participants[9]			Control			Treatment		
	n	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*
Academic SE[2]	805	23.8	3.7	269	23.8	3.8	300	23.8	3.9	743	23.9	4	241	23.9	4.4	303	23.5	3.9
Social SE[3]	806	26.2	4.9	270	26.1	5.1	300	26.5	5	743	26.4	5	241	26.3	5.1	303	26.6	5
Argumentation SE[4]	806	24.8	6.9	270	24.2	7.3	300	24.8	6.9	742	25	7.5	240	24.7	7.9	303	24.5	7.4
Nature of science[5]	801	41.2	8.2	269	40.2	8	298	40.8	8.1	738	41.1	8.6	236	41.6	8.9	303	39.7	8.2
Job knowledge[6]	803	6.7	3.1	269	6.2	3.2	299	6.9	2.9	742	6.8	3.5	240	6.8	3.5	303	6.4	3.5
O.M. exam[7]	837	5.5	2.7	269	5.2	3	337	5.4	2.5	754	7.1	3.4	267	7.2	3.5	303	7.2	3.2
Standards-based test[8]	—	—	—	—	—	—	—	—	—	670	13.1	4.7	251	13.4	5.3	260	13.1	4.5

[1]Means and standard deviations for sample and subsamples reported without listwise or pairwise deletion. [2]Sum of six Likert-type self-efficacy items (1 = *low*, 5 = *high*). [3]Sum of seven Likert-type self-efficacy items (1 = *low*, 5 = *high*). [4] Sum of nine self-efficacy Likert-type items (0 = *low*, 4 = *high*). [5]Sum of 17 Likert-type items, (0= *strongly disagree*, 4= *strongly agree*). [6]Sum of 17 dichotomous items. [7]Total number correct on a 16-item test. [8]Total number correct on a 24-item test. [9] The *n* for participants does not equal sum of control and treatment participants.

Table 8.
Comparison of DiSC 2005 Operation Montserrat Curriculum-oriented Exam Achievement Results with Previous Results Using the Full Pool of Operation Montserrat Achievement Test Items

	Pretest %	Posttest %	Gain %
DiSC 2005 Data[1]			
All	34	44	10
Control	33	45	12.5
Treatment	34	45	11.25
Challenger Center 2004 Data[2]			
AEP County	33	36	3
Non-AEP County	32	37	5

[1]Identical pre- and posttests with 16 items, drawn from the pool of Operation Montserrat Achievement Test items. [2]Results from a set of two parallel tests of 40 items each that, together, comprise the pool of Operation Montserrat Achievement Test items. Pre- and posttest data collected over 508 students in four states. AEP counties were targeted as impoverished counties. Non-AEP counties were not tracked for economic SES.

the Operation Montserrat unit and conducted the e-Mission. DiSC 2005 teachers were facilitated via telephone, the study web site, and e-mail support. They had never attended Operation Montserrat or e-Mission training. That DiSC gain percentages are similar (if not higher) to the AEP/non-AEP percentages suggests that facilitator and web site support was sufficient for DiSC 2005 teachers' implementation of the instructional unit paralleled the quality of the AEP/non-AEP teachers.

A repeated measures analysis of variance (ANOVA) with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (OM pretest and posttest) as the one within subjects variable was conducted to evaluate growth in achievement and to test for interaction between condition and gains because of instruction over time. Results revealed a significant main effect for instruction $F(1,503)=184.03$, $p<.001$, $\eta^2_{partial}=.27$. Students in both conditions increased their scores over time. This is a modest effect, on the borderline between a small and a medium effect. Cohen's explanation is helpful in interpreting effect sizes (EF, in this case calculated by the partial eta squared, $\eta^2_{partial}$):

> My intent was that the medium ES represents an effect likely to be visible to the naked eye of the careful observer. (It has been since noted in effect-size surveys that it approximates the average size of observed effects in various fields.) I set small ES to be noticeably smaller than medium but not so small as to be trivial, and I set large ES to be the same distance above medium as small was below it. Although the definitions were made subjectively, with some early minor adjustments, these conventions have been fixed since the 1977 edition of SPABS and have come into general use (p. 156).

Neither the main effect for DiSC usage nor the condition X instruction interaction were significant.

Standards-based test.

There is no significant difference in achievement on the standards-based test between the group that used the DiSC tool, $\overline{X} = 13.1$, $n=260$, and the placebo group, $\overline{X} = 13.4$, $n=251$. DiSC 2005 did not administer a pretest for the standards-based test, so no comparison can be made for the change in achievement over time.

Comparisons among the three assessments.

The three assessments are significantly correlated with each other (see Table 9). The medium, positive correlation ($r=.46$, $p<.01$) between the standards-based test (SBT) and the Operation Montserrat (OM) pretest is the lowest, followed by the correlation between the OM pre- and posttests ($r=.52$, $p<.01$). The correlation between the post OM exam and the STB test ($r=.69$, $p<.01$) is positive and high. This suggests that both OM and the SBT are capturing achievement gains, but the two tests are capturing different pictures of that achievement. Remember that the OM items were designed and selected specifically to align with the skills of the OM unit of instruction. If the SBT test had used standardized test items that were aligned to the instructional unit content and activities, its items would be described as "cherry-picked." Instead, the STB items were randomly selected from a pool of items that aligned with the state and national *standards* that aligned with the unit of instruction. Thus, the SBT provides an indication of how students might score on standardized test items covering Earth system science.

Table 9.
Correlations Among Assessments.[1, 2]

	Pre OM	Post OM	SBT
Pre-Operation Montserrat Curriculum-oriented Test (OM)	1		
Post-Operation Montserrat Curriculum-oriented Test (OM)	0.52**	1	
Standards-based Test (SBT)	0.46**	0.69**	1

[1] Listwise Deletion, n = 556
[2] two-tailed
** p=.01

On the average all students answered a greater percentage of STB questions correctly than OM questions correctly ($N_{all\ participants}$=850, 54.6%$_{SBT}$, 44.4%$_{OM}$), ($n_{DiSC\ tool}$ =303, 54.6%$_{SBT}$, 45.0%$_{OM}$), ($n_{placebo\ tool}$=251, 55.8%, 45.0%$_{OM}$). A difficulty index runs from 0 (so difficult no one answered correctly) to 100 (so easy that all test takers answered correctly). Comparison of difficulty indices (Figure 5 and Table 10 as well as Appendix F, Figure 1 and Figure 3) suggests that (a) the SBT has a greater percentage of very difficult items (difficulty index<30) and (b) the SBT has a larger number of items at the easier ranger (difficulty index ≥60). Computing the post-OM and SBT scores as percentages for just the participants included within the tool-usage subsample and running a paired, two-tailed t-test analysis yields that students scored significantly higher on the SBT ($\overline{X}_{SBT}$= 55%, SD = .20, n=485) than on the OM posttest ($\overline{X}_{postOM-}$= 45%, SD=.21 n=485), t(484)=13.7, p<.001, d=.54[4]. This is a medium effect size.

A repeated measures ANOVA with condition (DiSC tool or placebo tool) as the between-subjects variable and assessment instrument (OM posttest and SBT percentage score) as the within subjects variable was conducted to test the interaction between condition and assessment instrument. As was found with the paired t-test, the significant main effect for difference in achievement because of the test was significant with a modest effect size, F(1,483)=188.70, p<.001, $\eta^2_{partial}$=.28. The condition X test interaction was not significant, neither was the main effect for condition. There was a violation of Box's test of equality of covariance matrices within the repeated measures ANOVA. Conservative corrections offered by the SPSS software did not change the results.

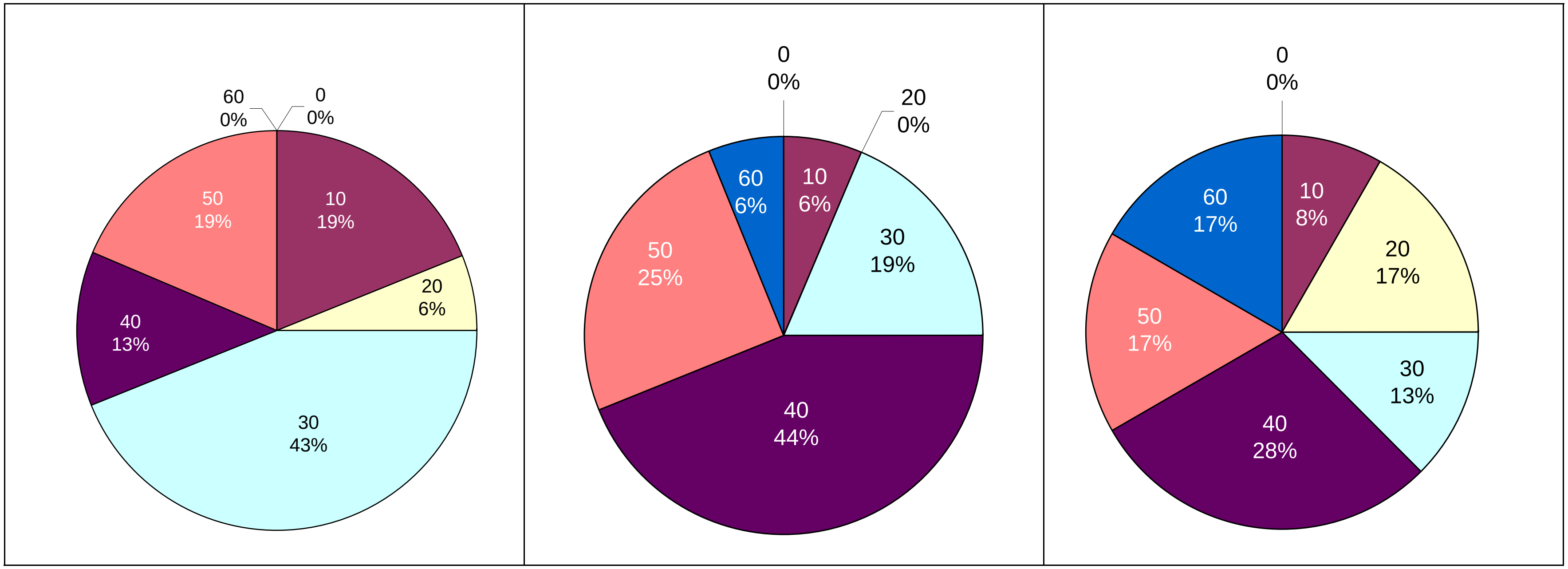
Self-efficacy, Nature of Science, and Job Knowledge

Pre-/Postcomparisons for the three self-efficacy constructs (academic achievement, social, and argumentation), nature of science, and job knowledge did not yield any significant differences because of the Operation Montserrat unit of instruction (see Table 11).

Self-efficacy.

Confining the analysis to the DiSC usage dataset, and running repeated measures analysis of variance (ANOVA) on the three self-efficacy constructs with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (each self-efficacy set of pre- and postdata) as the one within subjects variable yielded no significant main effects or interactions.

[4] Cohen's d is a measure of effect size: "negligible effect (>= -0.15 and <.15), small effect (>=.15 and <.40), medium effect (>=.40 and <.75), large effect (>=.75 and <1.10), very large effect (>=1.10 and <1.45), huge effect >1.45" (Thalheimer & Cook, 2002).



(a) OM pretest (b) OM posttest (c) Standards-based Test

Figure 5. Comparison of the Difficulty Indexes for the Three Academic Assessment Instruments. The pie graph labels indicate a difficulty level interval and the percentage of items for that test that scored within that interval. For the purposes of the pie chart, difficulty levels for items within each assessment were categorized into intervals by 10s (i.e., 0 to 9=0, 10 to 19=10, 20 to 29=20, 30 to 39 = 30, 40 to49 = 40, 50 to59 = 50, 60 to 69 = 60, etc.).

Table 10. Frequency of Items Within Each Interval: Comparison of the Difficulty Indexes for the three Academic Assessment Instruments.

Item Difficulty Level Interval	Operation Montserrat Pretest	Operation Montserrat Posttest	Standards-based Test
0	0	0	0
10	3	1	2
20	1	0	4
30	7	3	3
40	2	7	7
50	3	4	4
60	0	1	4

Table 11.
Paired *t*-test (Two-tailed) Comparisons for the Pre/Post Self-Efficacy, Nature of Science, and Job Knowledge Constructs.

	All Participants				DiSC Usage			
	N	*t(df)*	SD_{pre}	SD_{post}	*N*	*t(df)*	SD_{pre}	SD_{post}
Academic Self-efficacy[1]	630	-.089(629)	3.6	4	452	-.812(451)	3.7	4.1
Social Self-efficacy[2]	630	-1.1(629)	4.8	4.9	452	-.446(451)	4.1	4.9
Argumentation Self-efficacy[3]	629	-1.1(628)	6.7	7.4	451	-.272(450)	6.9	7.5
Nature of Science[4]	621	1.4(620)	8.3	8.8	445	1.3(444)	3	3.5
Job Knowledge[5]	626	-.15(625)	3.1	3.6	449	1.0(448)	8.2	8.7

[1]Sum of six Likert-type self-efficacy items (1= *low*, 5 = *high*). [2]Sum of seven Likert-type self-efficacy items (1 = *low*, 5 = *high*). [3] Sum of nine self-efficacy Likert-type items (0 = *low*, 4 = *high*). [4]Sum of 17 Likert-type items, (0 = *strongly disagree*, 4 = *strongly agree*). [5]Sum of 17 dichotomous items.
Note. None of the results were significant at the 0.05 level.

Nature of science.

A repeated measures analysis of variance (ANOVA) with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (pre- and postdata) as the one within subjects variable for nature of science yielded no significant main effect for time but a significant main effect for tool usage, $F(1, 443)=4.0$, $p<.05$, $\eta^2_{partial}=.009$. This is a negligible effect that must be interpreted in conjunction with the interaction between tool usage and instruction, $F(1,443)=5.2$, $p<.05$, $\eta^2_{partial}=.012$ (a small effect, Lipsey, 1990), ($\overline{X}_{\text{pre: control}}= 41.5$, $\overline{X}_{\text{pre: treatment}}= 41.1$, $\overline{X}_{\text{post: control}}= 42.0$, $\overline{X}_{\text{post: treatment}}= 39.7$). On the average and before instruction, learners in both the control and treatment group were relatively close in self-reports of the nature of science (about 41.25 points on a 68-point scale). During the course of instruction and on the average, the control group (students using the placebo tool) modified their mental model of the nature of science to a more dynamic view of science as a dynamic, culturally-mediated, social practice based upon observation, theory, reasons, and evidence. The change was small, about 1/2 unit. The treatment group (the learners who used the DiSC tool) moved, on the average, a bit toward the other side of the scale (a more static view of science as individual pursuit with a stable knowledge base that never changes). The change was not large, about 1.5 units within a scale of 68 units.

Job knowledge.

A repeated measures analysis of variance (ANOVA) with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (pre and post data) as the one within subjects variable for job knowledge yielded no significant main effect for time or tool usage, but a significant disordinal interaction[5] between the two main effects, $F(1, 447)=8.1$, $p<.01$, $\eta^2_{partial}=.018$. This is a small effect. On the average and before instruction, learners in both the control ($\overline{X} = 6.5$) and treatment group ($\overline{X} = 7.1$) were relatively close in self-reports of the nature of science (about 6.8 points on a 17-point scale). During the course of instruction and on the average, the control group (students using the placebo tool) modified their mental model of the nature of science to a more dynamic view of science as a dynamic, culturally mediated, social practice based upon observation, theory, reasons, and evidence. The average score for the students using the placebo tool gained 0.4

[5] A disordinal interaction is one in which lines drawn through the pre/post scores for each of the variables intersect within the interval under consideration (Pedhazur, 1997).

points and the average score for the students using the DiSC tool dropped 0.7 points. The overall average score across conditions is about 39 percent of the total available. This means that after exposure to the Operation Montserrat unit of instruction, students could correctly answer 7 of the 17 job knowledge questions about careers directly related to the science practice they had studied within the Operation Montserrat unit of instruction and simulated during the live event.

Perceived Challenge/Skill During the Operation Montserrat Unit of Instruction

Results are presented first as an examination of frequencies at the experience level for the four combinations of skill and challenge that represent dimensions of experience. Then results are presented at the student level, for the data aggregated by period for each student. Student-level analysis used a calculated variable, the geometric mean of aggregated skill and aggregated challenge, to measure perceived challenge/skill for each of the five periods.

Discovery motivates design modification.

DiSC 2005 data collection had intended to group ESM 30 within wave 4, but also to analyze the e-Mission ESM data separately to investigate characteristics of the e-Mission experience. COTF made a discovery during the analysis of the DiSC 2005 data that required a modification in the design of DiSC 2005 data analysis. DiSC had collected ESM data at the student level in four waves (baseline, wave 2, wave 3, and a final period) over the course of the unit. The e-Mission was the last data point within the final period, and COTF was also interested in the states and quality of experience during this live simulation. Therefore, COTF partitioned the final period into wave 4 (all final ESM data EXCEPT the e-Mission) and e-Mission (the final ESM collection which occurred 55 minutes into the e-Mission experience about halfway through the two-hour online adventure). Learners had spent four weeks engaged with Earth systems and related job knowledge content as well as argumentation concept and procedures. The e-Mission was designed as a transfer task during which learners applied what they had learned during the preceding four weeks toward solving a virtual crisis situation. The instruction supplied students with skills, and the e-Mission supplied the challenge. As this report will indicate in the following sections, learners reported significantly higher levels of skills and challenges during the e-Mission live simulation than during the preceding four weeks of classroom unit instruction. Therefore, it is misleading to create mean composites composed of ESM data from the wave 4 experiences and the e-Mission experience. The e-Mission ESM data measures student perceptions at the experience level, as do baseline and the other four waves. As argued within the assumptions subsection of this report (see Methods section), because all e-Mission data was collected at 55 minutes into the mission—the same time for all participants—this analysis assumes it is a valid indicator of experience at the midway point during the e-Mission. An ESM period normally samples participants' experiences over a wide range of daily activities. Flow is normally an aggregated mean, and five to eight to more experiences are aggregated within a period. This increases stability of the measure. During DiSC 2005 the only e-Mission experience sampled was at e-Mission time = 55. This makes the stability of the measure questionable. However, all participants were sampled at the same e-Mission minute, and those responses are aggregated over the entire tool usage dataset. This report is based upon the argument that it is valid to consider this one ESM sampling as a measure of e-Mission dimensions and quality of experience at time = 55 minutes.

Challenge/Skill Combination Trends at the Response Level

Recall, each individual's skill and challenge raw scores had been calculated via person-level[6] *z* scores. Then four dummy variables had been calculated, one for each state of experience; that is, each of the four combinations of skill and challenge (i.e., high challenge/skill, high challenge/low skill, low challenge/high skill, and low challenge/skill) using the methods discussed above[7]. Crosstabs were run for each dummy variable and each period (the baseline, each of the three waves, and the e-Mission) across the two conditions. The results were used to examine the proportion responses within each condition and period that represented one of the states of experiences (see Figures 6 – 9 and Table 11). Remember, results for each state of experience are the aggregate across the tool usage subsample of each individual participant's responses standardized in relation to all the ESM response recorded by that individual.

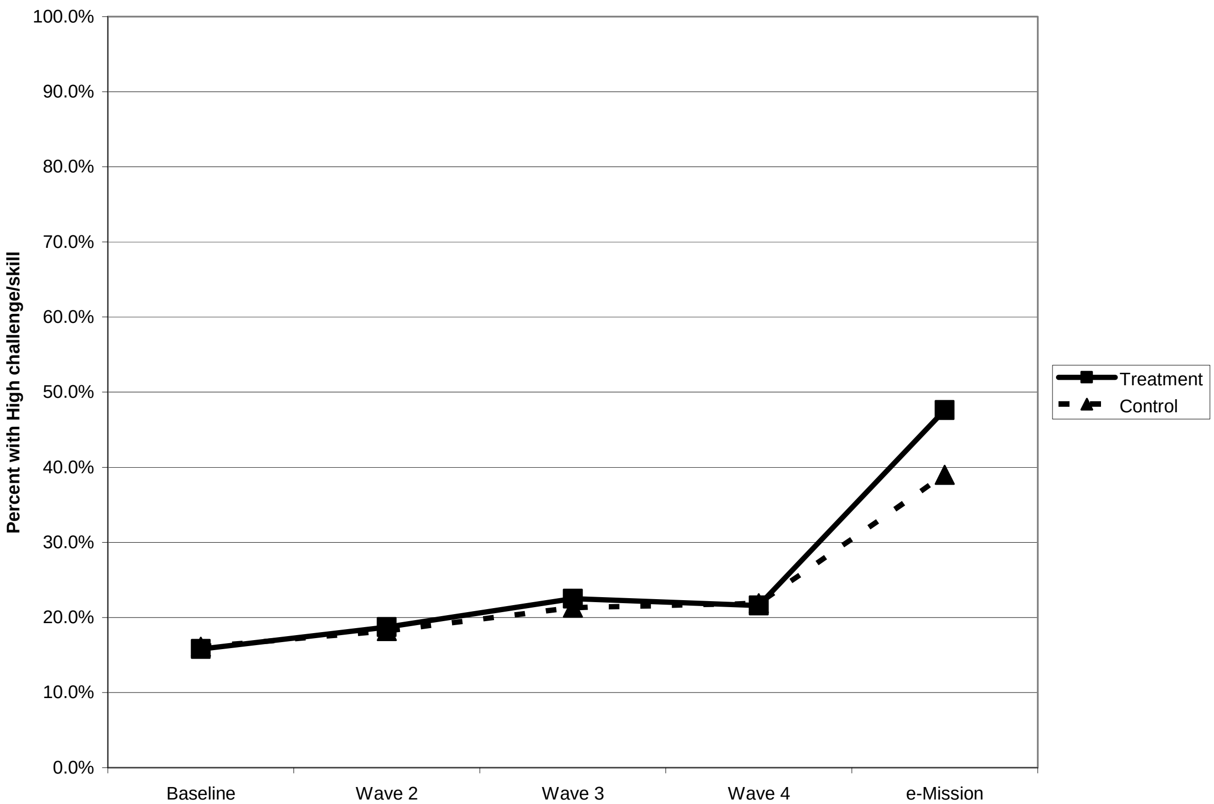


Figure 6. Experience-level Reports of High Challenge/Skill as a Percentage of the Total Challenge/Skill Reports for Each Condition by Period.
Note: Treatment = DiSC tool usage, Control = placebo tool usage. If either or both raw skill and challenge was a missing value, the dummy variable was computed as a missing value. Judgments were made on a 9-point scale (1= *low* , 9 = *high*).

---

[6] Within SPSS this can be accomplished by parsing the data by individual using the SPSS "split file" command and then calculating *z* scores. Each standardized variable response for an individual has then been standardized in relation to all the ESM responses for that item recorded by that individual. This is what is meant by "person-level" *z* scores. Thus, results for each state of experience are calculated from scores standardized at the person level.
[7] Thus, results for each state of experience are calculated from scores standardized at the person level.

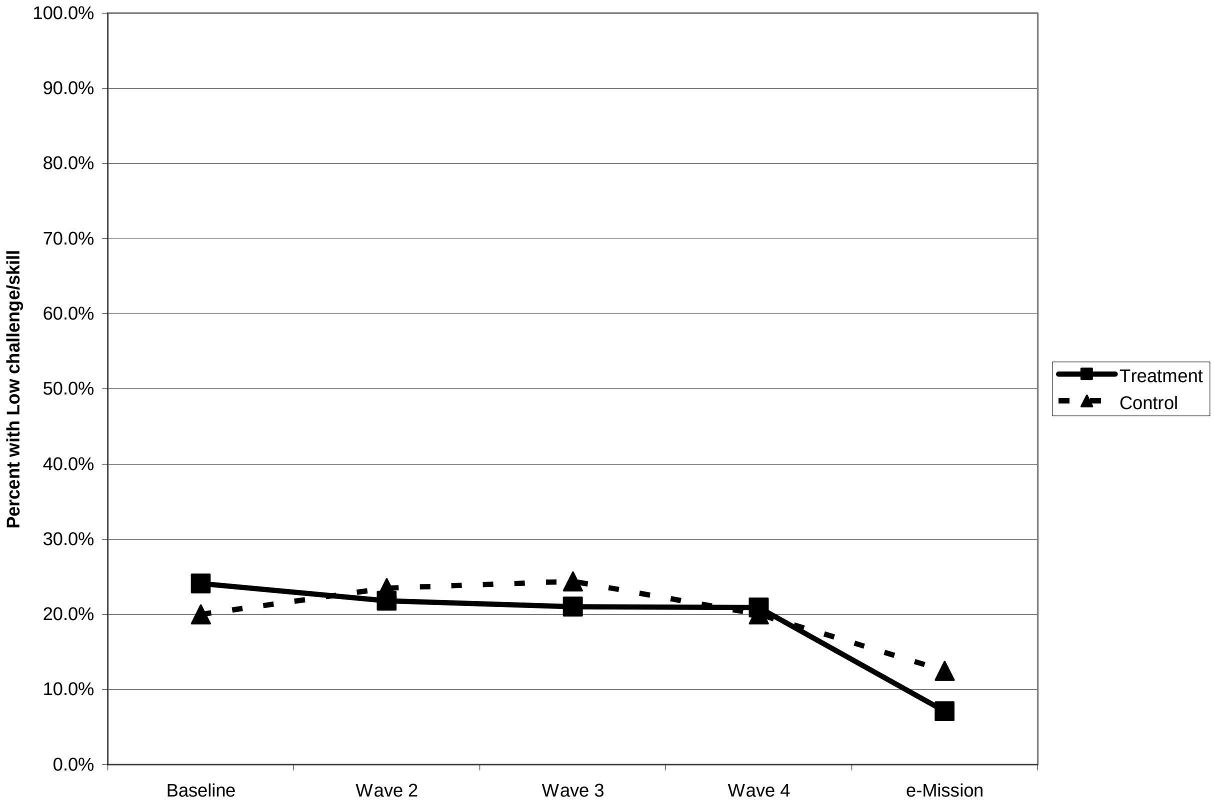


Figure 7. Experience-level Reports of Low Challenge/Skill as a Percentage of the Total Challenge/Skill Reports for Each Condition by Period.
Note: Treatment = DiSC tool usage, Control = placebo tool usage. If either or both raw skill and challenge was a missing value, the dummy variable was computed as a missing value. Judgments were made on a 9-point scale (1= *low* , 9 = *high*).

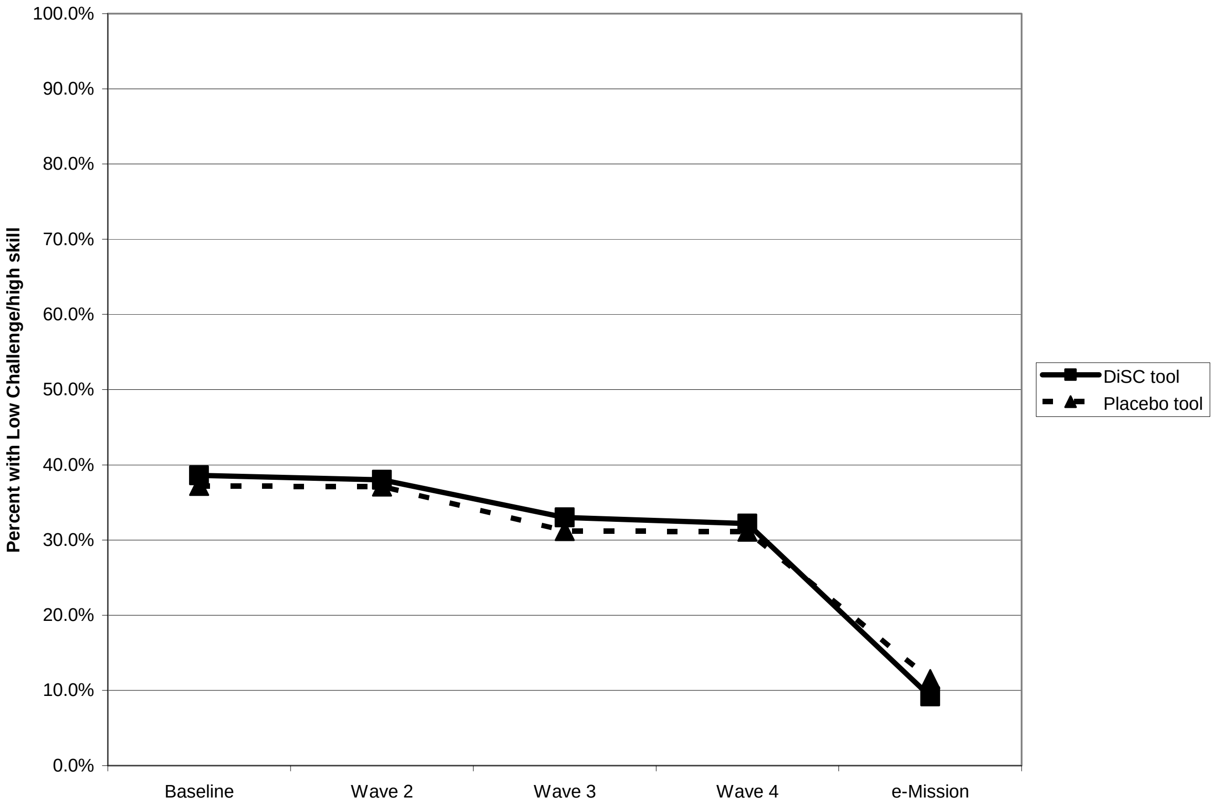


Figure 8. Experience-level Reports of Low Challenge/High Skill as a Percentage of the Total Challenge/Skill Reports for Each Condition by Period.
Note: Treatment = DiSC tool usage, Control = placebo tool usage. If either or both raw skill and challenge was a missing value, the dummy variable was computed as a missing value. Judgments were made on a 9-point scale (1= *low* , 9 = *high*).

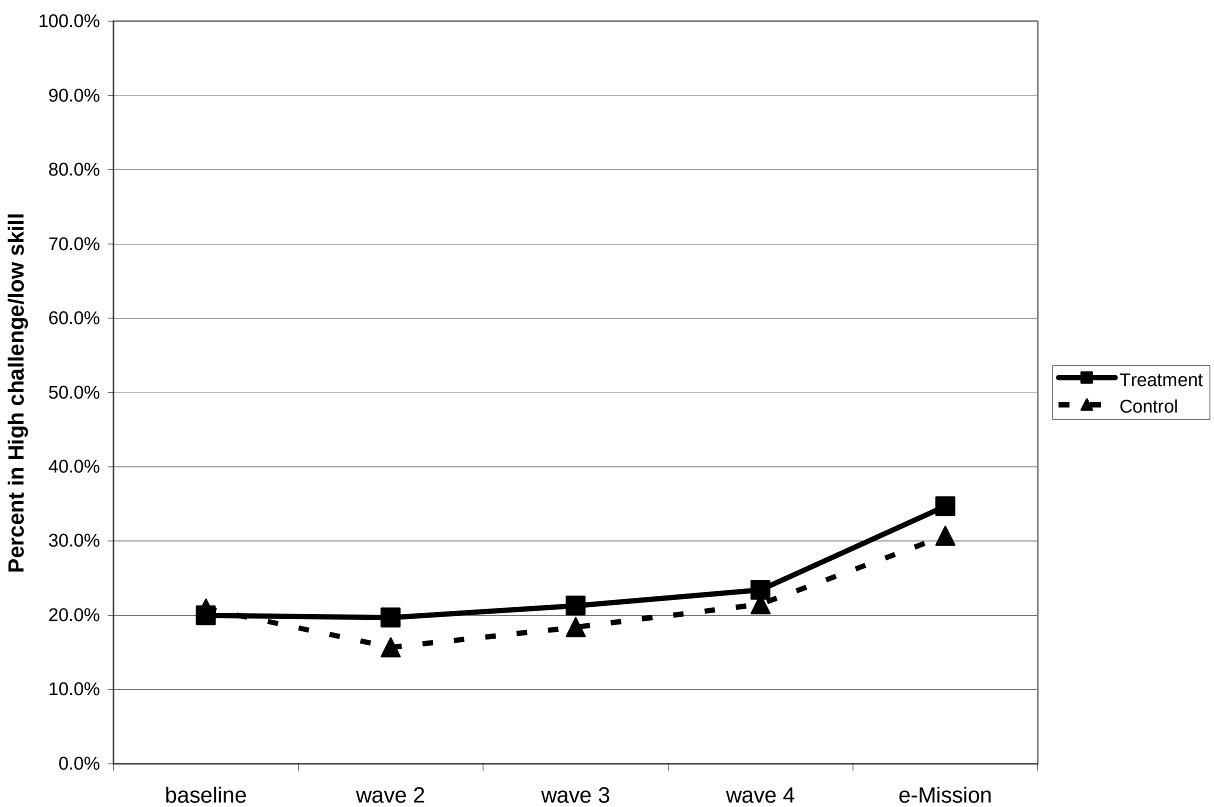


Figure 9. Experience-level Reports of High Challenge/Low Skill as a Percentage of the Total Challenge/Skill Reports for Each Condition by Period.
Note: Treatment = DiSC tool usage, Control = placebo tool usage. If either or both raw skill and challenge was a missing value, the dummy variable was computed as a missing value. Judgments were made on a 9-point scale (1= *low* , 9 = *high*).

Table 11.
Percentage of Responses in State[1] for Each Condition[2] by Period (at the Experience Level)

	Baseline	Wave 2	Wave 3	Wave 4	e-Mission
Challenge/Skill					
Treatment	15.8%	18.7%	22.5%	21.6%	47.6%
Control	15.8%	17.9%	21.3%	21.4%	39.0%
Low challenge/Skill					
Treatment	24.1%	21.8%	21.0%	20.9%	7.1%
Control	20.3%	23.5%	24.2%	20.1%	12.5%
High challenge/Low skill					
Treatment	20.0%	19.7%	21.3%	23.4%	34.7%
Control	20.8%	15.5%	18.3%	21.5%	30.7%
Low challenge/High skill					
Treatment	38.6%	38.0%	33.0%	32.2%	9.2%
Control	37.3%	37.8%	31.4%	31.8%	11.5%
Number of Responses					
Treatment	1612	2719	2535	1670	294
Control	1208	2092	2153	1399	287
Total	2820	4811	4688	3069	581

[1] If either or both raw skill and challenge was a missing value, the dummy variable was computed as a missing value. [2]Treatment = DiSC tool usage, Control = placebo tool usage.
Note: Judgments were made on a 9-point scale (1= *low* , 9 = *high*).

High challenge/skill.

Overall, there was an increase in high challenge/skill reported (see Figure 6 and Table 11) from a baseline average of 16 percent to the e-Mission of average of 43 percent ($\text{increase}_{\text{placebo tool}}$=23%, $\text{increase}_{\text{DiSC tool}}$=32%). Increase for the condition using the DiSC tool was 9 percent greater than the increase for the group using the placebo tool. During the e-Mission about 9 percent more of the DiSC tool users reported high challenge/skill than the placebo tool users. The sharp spike that occurred from wave four to e-Mission suggests the e-Mission live simulation is a different experience than the other unit activities (which are more like the typical classroom learning environment). The larger percentage gains from baseline to e-Mission, and the larger report of high challenge/skill by the DiSC group suggests that DiSC usage during instruction increased perceived high challenge/skill during the e-Mission. There was also a slight, but higher report of high challenge/skill by the DiSC users during wave 3.

Low challenge/skill.

Experience-level reports of Low challenge/skill (see Figure 7 and Table 11) decreased from baseline ($24.1\%_{\text{DiSC tool}}$, $20.0\%_{\text{placebo tool}}$) to e-Mission ($7.1\%_{\text{DiSC tool}}$, $12.5\%_{\text{placebo tool}}$), from an average of 22 percent to an average of 10 percent (see Figure 7). The e-Mission percentage is lower than any of the preceding periods. The DiSC group perceived low challenge/skill more frequently during baseline than the placebo group. However, DiSC users logged fewer reports of low challenge/skill at e-Mission time = 55 than did the placebo tool users. This suggests that (a) more DiSC learners were engaged during the e-Mission and (b) the DiSC tool prepared learners to be more engaged during the e-Mission. The placebo tool group also recorded more responses in this state during wave 3. Although the difference is not as dramatic as the overall e-Mission drop, it is notable that the DiSC tool users were higher than placebo at baseline and lower at wave 3 (about 3 percent fewer responses during this period).

Low challenge/High skill.

The near-overlap of the two lines across all five periods indicates that reports of low challenge/high skill were very consistent across both conditions for this state (see Figure 8). Reports of relaxation decreased from about 38 percent during baseline to about 31.5 percent during wave 4. At e-Mission time = 55, far fewer participants reported relaxed states in which skills were high and challenge low. Across both conditions only about 10 percent of responses indicated this state. This is a decrease of about 22 percent. Coupled with results for the other states, this indicates that the e-Mission, at least at time = 55, was challenging and the majority of students were not simply coasting through the e-Mission experience.

High challenge/Low skill.

Notice (see Figure 9 and Table 11) about an equal proportion of DiSC usage students reported high challenge/low skill at baseline, followed by a larger proportion of DiSC usage students reporting high challenge/low skill in every subsequent period. This indicates that placebo tool users perceived less challenge during waves 2, 3, and 4 and during the e-Mission than did the DiSC tool users.

Aggregates at the student/period level

Given the assumptions, the data can also be examined using inferential statistics by calculating aggregates for each variable of interest by period for each individual student[8]. The challenge/skill student-level composite was calculated for each period. Then tool usage dataset was filtered for participants with three or more baseline responses.

Challenge/Skill changes from baseline to e-Mission.

A repeated measures analysis of variance (ANOVA) with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (pre- and postdata) as the one within subjects variable for challenge/skill yielded a significant main effect for instruction $F(1, 486)=166.91, p<.001, \eta^2_{partial}=.26$ (see Figure 10). The main effect for condition was not significant but approached significance at $F(1,486)=3.9, p=0.0501$. Mean values ($\overline{X}_{pre: control}= 4.53$, $\overline{X}_{pre: treatment}= 4.46$, $\overline{X}_{post: control}= 5.54$ $\overline{X}_{post: treatment}= 6.10$, $n_{control}=210$, $n_{treatment}=278$) indicate that DiSC users recorded higher challenge/skill at e-Mission time = 55 minutes, with a significant but small disordinal interaction[9] between the two main effects, $F(1, 486)=9.71, p<.01, \eta^2_{partial}=.02$ (see Figure 10). Before instruction the DiSC group indicated lower challenge/skill than the learners in the control group. Based upon the 9-point ESM scale, the differences from baseline to e-Mission represent a 19 percent gain for the DiSC condition ($gain_{DiSC}=1.7$) and a 12 percent gain for the control group ($gain_{placebo}=1.1$).

---

[8] Use the SPSS "select cases" menu to select a period of interest. Then use the "aggregate" menu to specify the variables for aggregation. Select the "number of cases" option to record the number of ESM observations included within the aggregation for each individual.

[9] A disordinal interaction is one in which lines drawn through the pre/post scores for each of the variables intersect within the interval under consideration (Pedhazur, 1997).

Perception of skills and challenges are higher during the e-Mission, and there is an interaction between perceived skills and challenges and treatment: DiSC tool users, on the average, reported higher skills and challenges. Although significant, this is a small effect.

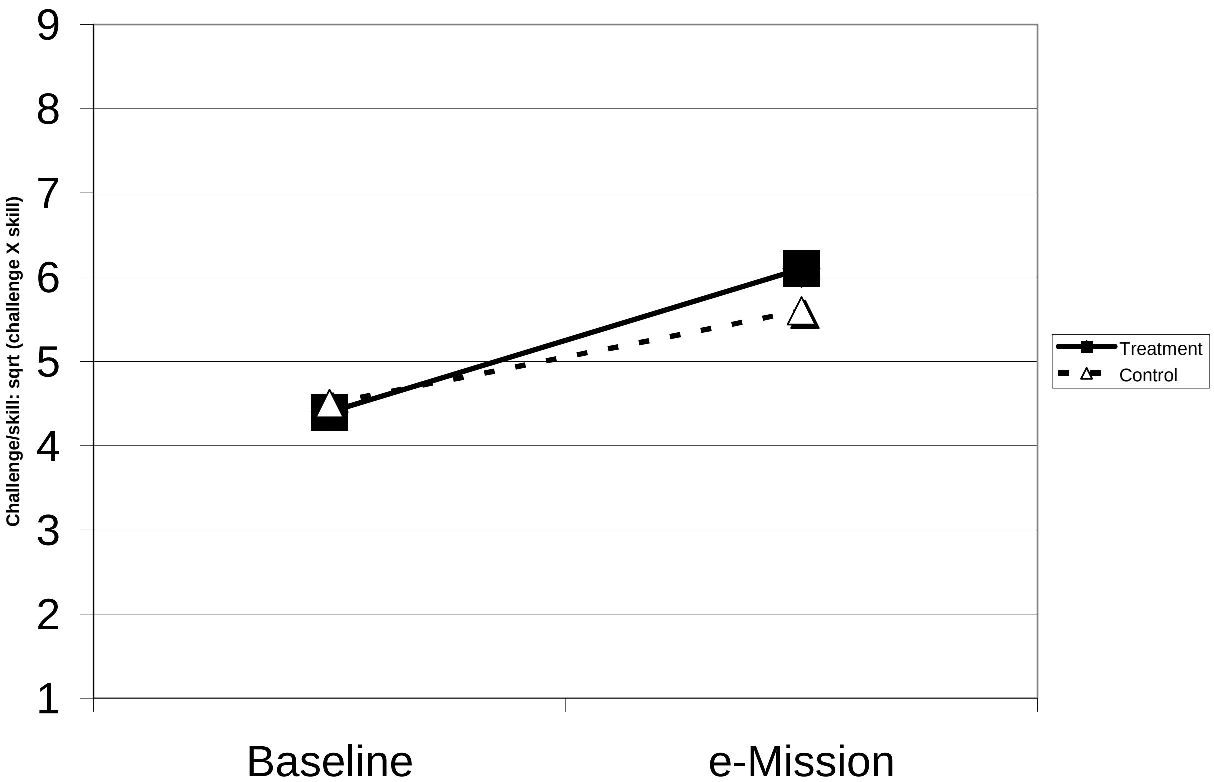


Figure 10. DiSC increases Challenge/Skill in Transfer to Live Simulation (e-Mission): The Interaction Between Instruction and Condition.
Note: Judgments were made on a 9-point scale (1= *low* , 9 = *high*).

Challenge/Skill changes across five periods.

The tool usage dataset was filtered for minimal response criteria (baseline = >2 and waves 2-4 > 4). A repeated measures analysis of variance (ANOVA) with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (each of the five periods) as the one within subjects variable for challenge/skill[10] yielded a significant main effect for instruction $F(4, 276)=29.35, p<.001$, $\eta^2_{partial}=.30$ (see Table 12). The main effect for condition was not significant. The small, interaction between condition and instruction (Figure 11), $F(4, 276)=2.53, p<.05$, $\eta^2_{partial}=.04$, parallels the frequency responses over dummy variables for the five periods reported above (see Figure 6).

[10] A Box's Test of Equality of Covariance Matrices indicated that the observed covariance matrices of the dependent variables were not equal across groups. Corrections by SPSS software did not change the degrees of freedom or the results.

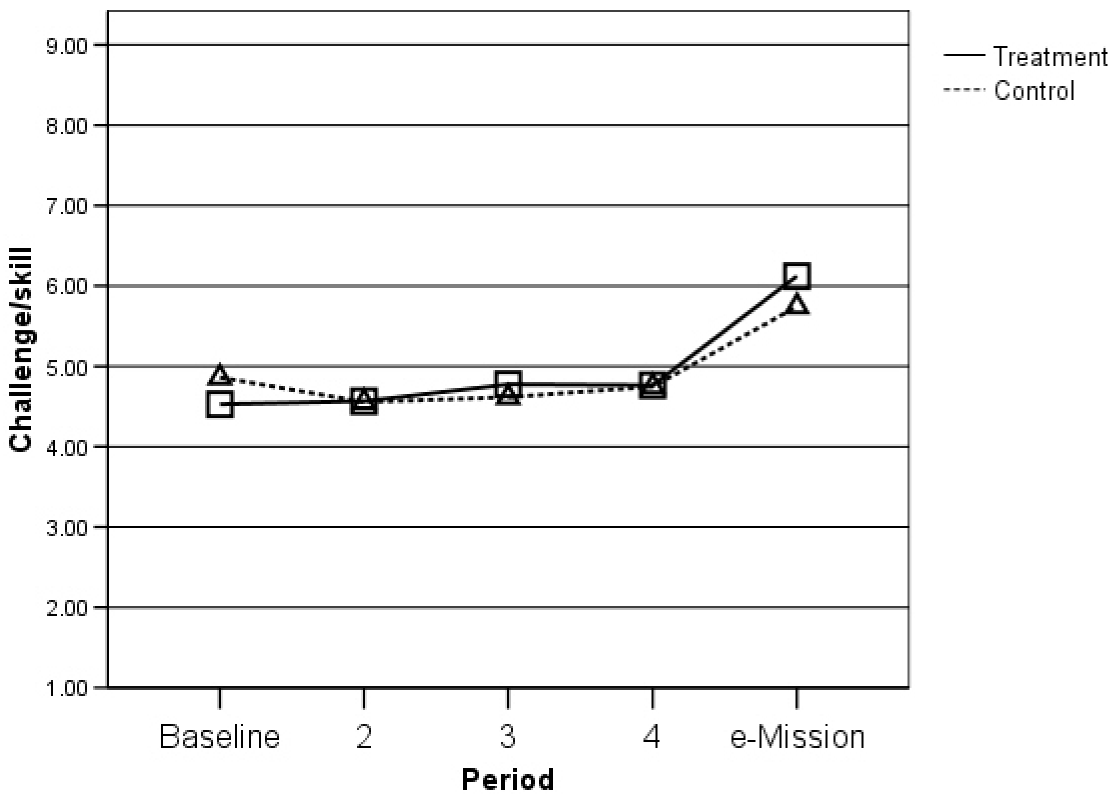


Figure 11. Changes in Mean Challenge/Skill Over the Five Periods of the Operation Montserrat Unit.
Note: Judgments were made on a 9-point scale (1= *low* , 9 = *high*). Listwise deletion: $n_{\text{control}} = 98$, $n_{\text{treatment}} = 183$, $n_{total} = 281$. Challenge/skill $= (challenge * skill)^{1/2}$ of ESM responses aggregated at the student level by period. Dataset filtered for minimal response criteria (baseline = >2 and waves 2-4 > 4).

Table 12.
Means and Standard Deviations for Repeated Measures ANOVA by Condition with Instruction: Challenge/Skill score (Period)[1].

	$\overline{X}$	*SD*
Challenge/Skill 1 (baseline)		
Control[2]	4.86	1.33
Treatment[3]	4.53	1.28
Total[4]	4.64[a]	1.31
Challenge/Skill 2 (wave 2)		
Control[2]	4.56	1.83
Treatment[3]	4.57	1.51
Total[4]	4.56[a]	1.62
Challenge/Skill 3 (wave 3)		
Control[2]	4.62	1.80
Treatment[3]	4.78	1.60
Total[4]	4.72[a]	1.67
Challenge/Skill 4 (wave 4)		
Control[2]	4.75	1.85
Treatment[3]	4.77	1.64
Total[4]	4.76[a]	1.71
Challenge/Skill 5 (e-Mission)		
Control[2]	5.75	2.13
Treatment[3]	6.12	2.05
Total[4]	5.99	2.08

[1]Listwise deletion. [2]$n_{\text{control}} = 98$. [3]$n_{\text{treatment}} = 183$. [4]$n_{total} = 281$.
[a] Mean difference between e-Mission and each of the other periods is significant at the .05 level, with Bonferroni adjustment for multiple comparisons.

Note: Challenge/Skill $= (challenge * skill)^{1/2}$ of ESM responses aggregated at the student level by period. Judgments were made on a 9-point scale (1= *low* , 9 = *high*). Dataset filtered for minimal response criteria (baseline = >2 and waves 2-4 > 4).

Disaggregating by period and condition, paired *t*-tests illustrate that e-Mission challenge/skill gains are significant in comparison to each of the other four periods for both conditions (see Table 13). Mean differences are significant, even if the Bonferroni correction (a statistical adjustment for the multiple comparisons in which the alpha level is divided by the number of comparisons) is applied. The effect sizes for the control conditions are medium ($.40 \leq d < .75$). Treatment baseline and wave 2 results are large effects ($.75 \leq d < 1.10$). Treatment wave 3 and wave results are medium effect sizes.

This set of analyses provides evidence that student perceptions of challenge/skill level during the e-Mission live simulation are higher than those of any of the other periods. This is a substantive result. The analyses also suggest that the students who used the DiSC tool perceived higher levels of challenge/skill during the e-Mission than did students who used the placebo tool.

Table 13.
Paired *t*-Tests of Challenge/Skill by Period[1] with e-Mission Period.

	$\bar{X}_{difference}$	*SD*	*t(df)*[2]	Cohen's *d*
Challenge/Skill 1 (baseline)				
Control[2]	.88	2.11	4.14(97)**	0.5[⊥⊥]
Treatment[3]	1.59	2.20	9.8(182)**	0.94[⊥⊥⊥]
Challenge/Skill 2 (wave 2)				
Control[2]	1.19	2.09	5.64(97)**	0.6[⊥⊥]
Treatment[3]	1.56	2.02	10.42(182)	0.87[⊥⊥⊥]
Challenge/Skill 3 (wave 3)				
Control[2]	1.13	2.08	5.37(97)**	0.57[⊥⊥]
Treatment[3]	1.35	2.06	8.85(182)**	0.73[⊥⊥]
Challenge/Skill 4 (wave 4)				
Control[2]	1.00	1.84	5.37(97)**	0.5[⊥⊥]
Treatment[3]	1.36	2.05	8.97(182)**	0.73[⊥⊥]

[1]Listwise deletion. [2]$n_{control}$ = 98. [3]$n_{treatment}$ = 183.
[a] Mean difference between e-Mission and each of the other periods is significant at the .05 level, with Bonferroni adjustment for multiple comparisons.
** = p < .001.
[⊥⊥] medium effect size. [⊥⊥⊥] large effect size.

Note: Challenge/Skill $= (challenge * skill)^{1/2}$ for ESM responses aggregated at the student level by period. Judgments were made on a 9-point scale (1= *low* , 9 = *high*). Dataset filtered for minimal response criteria (baseline = >2 and waves 2-4 > 4).

Perceived e-Mission Challenge/skill and Academic Achievement.

Filtering the tool usage dataset for minimal response criteria (baseline = >2, waves 2-4 > 4), academic achievement on the standards-based test (SBT) and e-Mission perceived challenge/skill are correlated *r*=.208, *p*=<.01, *n* =225, but SBT and baseline perceived challenge/skill are not correlated, *r*=-.019, *ns*, *n* =191. Running a regression analysis with perceived academic achievement (STB) as the dependent variable and perceived e-Mission challenge/skill as predictor yields an R^2= .043, meaning that challenge/skill 5 predicts 4% of the variance in SBT achievement, *F*(1,223) =10.121, *p*=.002; *t(1)*=3.18, *b*=.465, *p* = .002. The beta coefficient (*b*) predicts that for every 1 unit of change in e-Mission Challenge/skill, the standards-based test score goes up about ½ point, or 2 percent.

Considering Context: Effects of Human Capital and Nesting Within Classrooms, Teachers, and Schools

Consideration of context led to an exploratory analysis of the interaction of parents' level of education, as an indicator of students' human capital, with tool usage condition as a predictor of change in challenge/skill over time. The motivation was to discover a trend in the data that could be used for disaggregation for context effects (Schoenfeld, 2006, p. 17). The fact that students were nested within classrooms, classrooms within participating teachers, and teachers within schools led to a preliminary use of linear mixed models (Luke, 2004; Singer, 1998; SPSS, 2005; UCLA Academic Technology Services, 2006), which would control for the effect of parents' education and confirm if any effects remained once the nested characteristics of the study were controlled. School-level SES was added to this analysis. This section presents scatter plots that illustrate the change over time in DiSC and placebo tool users' challenge/skill (from baseline challenge/skill to e-Mission challenge/skill), controlling for parents' level of education. It then presents a repeated measures ANOVA mixed design analysis of growth over time in challenge/skill, as predicted by tool condition with parents' level of education as covariate. The graph of the results is presented. Then a linear mixed models repeated measures analysis of the same data with the addition of school SES as a covariate is described and the output graph is presented. Detailed results of this analysis are not presented, but a graph of model predictions is compared to the ANOVA analysis predictions. These findings were used to motivate a disaggregation of the data into two groups by more or less parents' level of education for a final analysis using *t*-tests on each group.

Graphical representations of the relationship between human capital and e-Mission challenge/skill.

Distribution of parents' education level ($\sqrt{education_{father} * eduation_{mother}}$)[11]is quite similar across treatments (see Figure 12). Scatter plots of the relationship between parents' level of education and e-Mission challenge/skill illustrate some differences between the two conditions (see Figure 13):

- There is a higher concentration of students in the *control* version with *lower* e-Mission challenge/skill when parents' education level is at the *low* end of the scale.
- There is a higher concentration of students in the *treatment* version with *higher* e-Mission challenge/skill when parents' education level is at the *low* end of the scale.
- There is a higher concentration of students in the *treatment* version with *lower* e-Mission challenge/skill when parents' education level is at the *high* end of the scale.

The scatterplot and regression of e-Mission challenge/skill on parents' level of education suggest that, for the treatment group scaffolded by the DiSC tool, e-Mission challenge/skill was independent of parents' level of education $F(1, 247)=.073$, *ns,* $R^2<.001$. The e-Mission challenge/skill *was* predicted by parents' level of education for the control group, $F(1, 215)=12.61$, $p<.001$, $t(1)<.001$, $b=.58$. The model for the control group predicts that for every 1 unit increase in parents' level of education (scale runs from 1-5), e-Mission challenge/skill increases about 0.6 challenge/skill units (scale runs from 1-9).

---

[11] As previously stated, scale translations for composite parents' education level (pairs are not ordered) are 1=(< high school ,< high school), 1.41=(< high school, high school), 1.73=(< high school, college), 2=(high school, high school) or (< high school, master's), 2.24=(< high school, PhD/MD), 2.45=(high school, college), 2.83=(high school, master's), 3=(college, college), 3.16=(high school, Ph.D./M.D.), 3.46=(college, master's), 3.87=(college, Ph.D./M.D.), 4=(master's, master's), 4.47=master's, Ph.D./M.D.), 5=(Ph.D./M.D., Ph.D./M.D.). Composite calculations drop cases with any missing data.

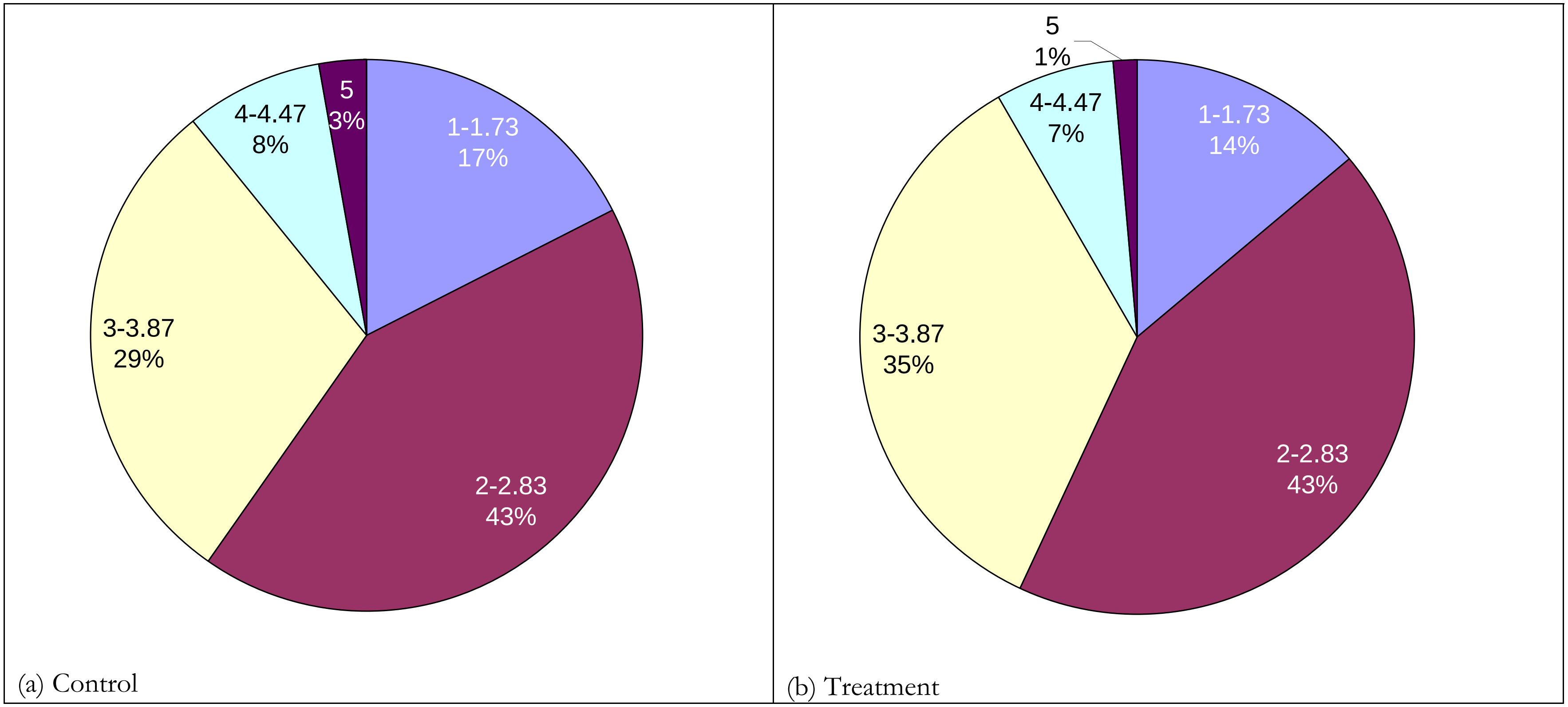
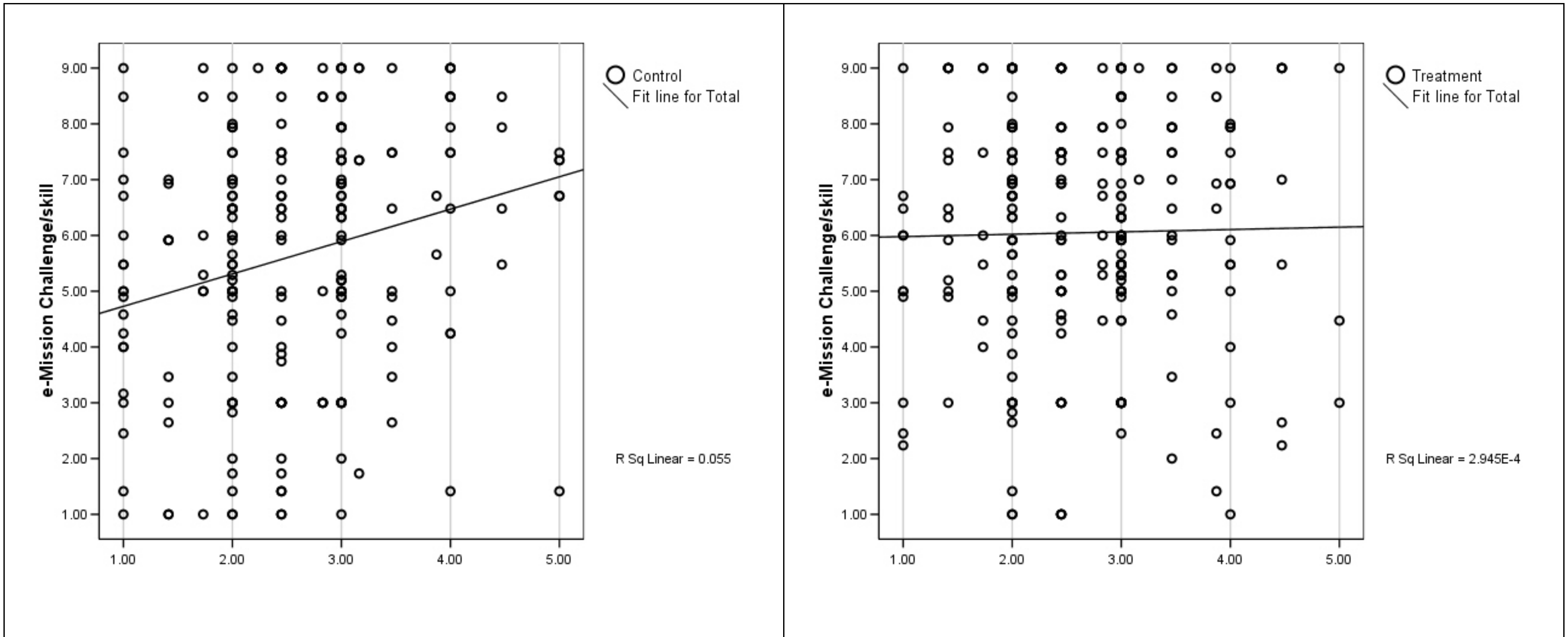


Figure 12. Comparison Between Control (Placebo) and Treatment (DiSC) Subsamples on Distribution of Parents' Education Level Using the Tool Usage Dataset.

Note: The pie graph labels indicate the parents' level of education category interval (lower number) and the proportion of subjects in that category (lower number). Parents' level of education is a composite, calculated using the geometric mean ($\sqrt{education_{father} * eduation_{mother}}$) of student-reported parent education levels. Scale translations for composite parents' level of education (pairs are not ordered): 1=(< high school ,< high school), 1.41=(< high school, high school), 1.73=(< high school, college), 2=(high school, high school) or (< high school, master's), 2.24=(< high school, Ph.D./M.D.), 2.45=(high school, college), 2.83=(high school, master's), 3=(college, college), 3.16=(high school, Ph.D./M.D.), 3.46=(college, master's), 3.87=(college, Ph.D./M.D.), 4=(master's, master's), 4.47=master's, Ph.D./M.D.), 5=(Ph.D./M.D., Ph.D./M.D.). Composite calculations drop cases with any missing data.



(a) Control

(b) Treatment

Figure 12. Scatterplots of Parents' Level of Education for Control and Treatment Subsamples Using the Tool Usage Dataset.
Note: $n_{control}$=217, $n_{treatment}$=249. Parents' level of education is a composite, calculated using the geometric mean ($\sqrt{education_{father} * eduation_{mother}}$) of student-reported parent education levels. Scale translations for composite parents' education level (pairs are not ordered): 1=(< high school ,< high school), 1.41=(< high school, high school), 1.73=(< high school, college), 2=(high school, high school) or (< high school, master's), 2.24=(< high school, Ph.D./M.D.), 2.45=(high school, college), 2.83=(high school, master's), 3=(college, college), 3.16=(high school, Ph.D./M.D.), 3.46=(college, master's), 3.87=(college, Ph.D./M.D.), 4=(master's, master's), 4.47=master's, Ph.D./M.D.), 5=(Ph.D./M.D., Ph.D./M.D.). Composite calculations drop cases with any missing data.

The interaction between human capital and challenge/skill.

A repeated measures analysis of variance (ANOVA) with condition (DiSC tool or placebo tool) as the between-subjects variable and instruction (baseline or e-mission challenge/skill) as the one within subjects variable for challenge/skill, and parents' level of education as the within subjects covariate was conducted to study the interaction between parents' level of education and changes in challenge/skill. The analysis yielded a significant main effect for instruction, $F(1, 459)=6.23, p<.05$, $\eta^2_{partial}=.01$. The main effect for condition was also significant but very weak, $F(1, 459)=4.28, p<.05$, $\eta^2_{partial}=.009$. The covariate made a significant contribution to the model, $F(1,486)=7.29, p=0.016$. The interaction between condition and parents' education level was not significant.

Consistent with the analyses above, the interaction between condition and instruction was significant and small, $F(1, 459)=8.15, p<.01$, $\eta^2_{partial}=.017$. As suggested by the scatterplots, the three-way interaction between condition, instruction, and parents' level of education was also significant, $F(1, 459)=4.48, p<.05$, $\eta^2_{partial}=.01$.

The model was used to estimate means for challenge/skill at 0.25 intervals of parents' level of education (see Figure 13):

- Estimated baseline means show little difference between DiSC and Placebo groups across all levels of parental education, with baseline challenge/skill slightly higher for the students using the placebo tool.
- Estimated baseline means show a slight increase in baseline challenge/skill across the range of parents' level of education.
- All estimated e-Mission challenge/skill means are higher than baseline challenge/skill means.
- Estimated e-Mission challenge/skill for the DiSC usage students is stable across all parents' education levels. In other words, it is independent of parent's level of education, and the best predictor is its mean value (6.05).
- Estimated e-Mission challenge/skill means for the placebo usage students is slightly higher than baseline for students whose composite parents' education level is close to 1.
- Estimated e-Mission challenge/skill means for the placebo usage students increase as parents' education level increases.
- The model estimates that control and treatment e-Mission challenge/skill intersect near a parents' education level of 3.5 (3.46 = a college, master's combination).
- The difference between estimated control and treatment e-Mission challenge/skill is larger at the lower end of the parents' level of education scale.

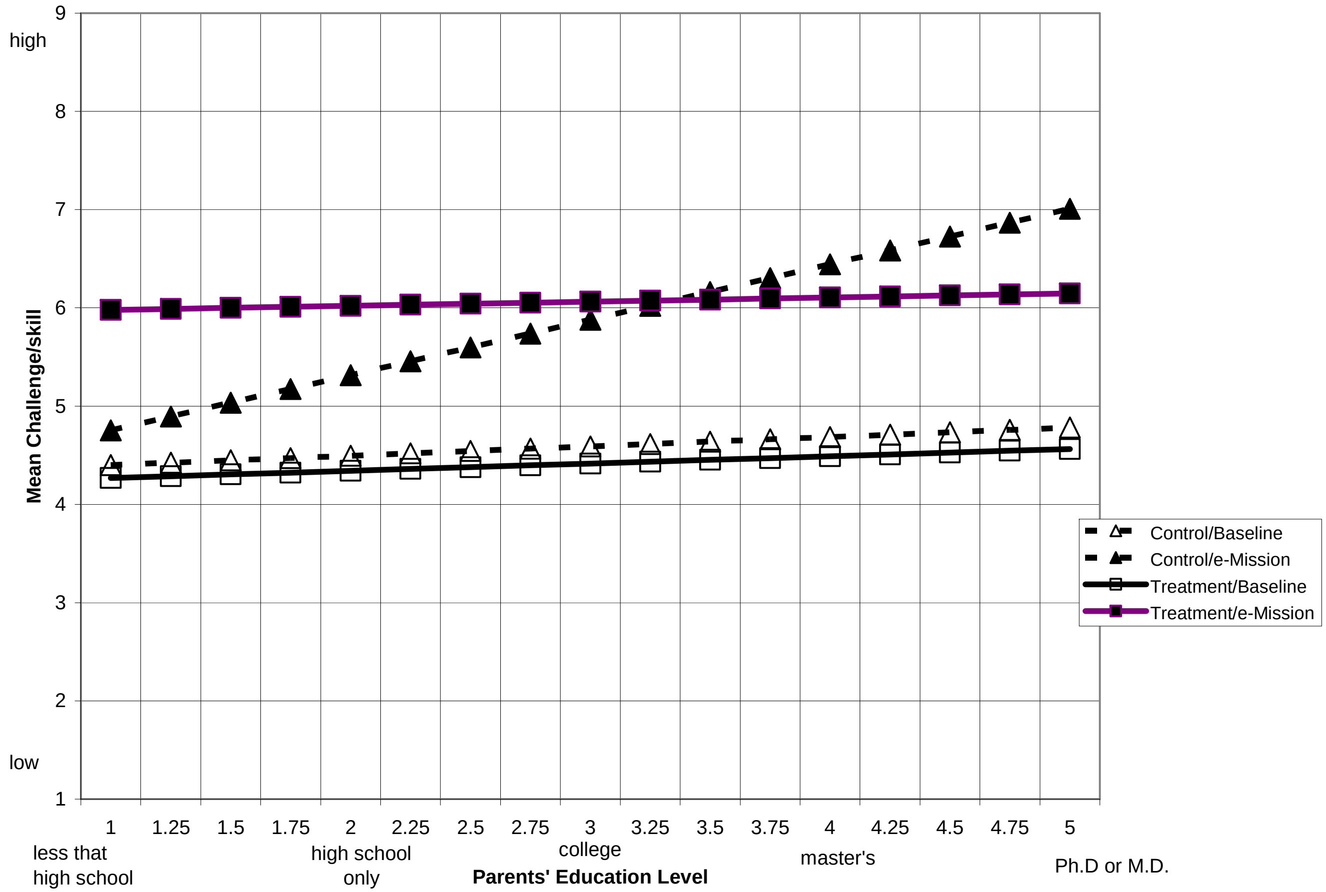


Figure 13. Four Regression Lines for Mean Perception of Challenge/Skill Against Parents' Level of Education, One for Each Cell of the Design.

Note: Scale translations for composite parents' level of education (pairs are not ordered): 1=(< high school ,< high school), 1.41=(< high school, high school), 1.73=(< high school, college), 2=(high school, high school) or (< high school, master's), 2.24=(< high school, Ph.D./M.D.), 2.45=(high school, college), 2.83=(high school, master's), 3=(college, college), 3.16=(high school, Ph.D./M.D.), 3.46=(college, master's), 3.87=(college, Ph.D./M.D.), 4=(master's, master's), 4.47=master's, Ph.D./M.D.), 5=(Ph.D./M.D., Ph.D./M.D.). Composite calculations (geometric mean of mother and father's education level) drop cases with any missing data.

Controlling for nesting and school level effects[12].

It is possible that three-way interaction between condition (DiSC or placebo tool), challenge/skill, and parents' level of education is a function of either school-level socioeconomic characteristics or an artifact of students sharing the same class, teacher, or school. An exploratory analysis using SPSS mixed models was run using a backward stepwise approach to build a linear growth model containing *instruction* (the repeated variable for baseline challenge/skill and e-Mission challenge/skill), and random intercepts at each level of the hierarchy (class, teacher, and school) above the level of the individual student. Other variables in the model were parents' level of education (Parent_Ed), School's SES (School_SES):

$$Y_{ijklst} = \mu + \alpha_s + \gamma_t + \alpha\gamma_{st} + \beta_1\text{*School\_SES} + \alpha\beta_{s1}\text{*School\_SES} + \gamma\beta_{t1}\text{*School\_SES} +$$

$$\alpha\gamma\beta_{st1}\text{*School\_SES} + \beta_2\text{*Parent\_Ed} + \alpha\beta_{s2}\text{*Parent\_Ed} + \gamma\beta_{t2}\text{*Parent\_Ed} + \alpha\gamma\beta_{st2}\text{*Parent\_Ed}$$

$$+ \beta_{12}\text{*School\_SES*Parent\_Ed} + \alpha\beta_{s12}\text{*School\_SES*Parent\_Ed} +$$

$$\gamma\beta_{t12}\text{*School\_SES*Parent\_Ed} + \alpha\gamma\beta_{st12}\text{*School\_SES*Parent\_Ed} + \mu_l + \mu_{kl} + \mu_{jkl} + \varepsilon_{ijklst}$$

Where i indexes subjects, j indexes classes, k indexes teachers, l indexes schools, s indexes instruction, and t indexes condition.

Model building details and full details of the mixed model analysis are not provided. The final model was:

$$Y_{ijklst} = \mu + \alpha_s + \gamma_t + \alpha\gamma_{st} + \beta_1\text{* School\_SES} + \alpha\beta_{s1}\text{* School\_SES} +$$

$$\beta_2\text{* Parent\_Ed} + \gamma\beta_{t2}\text{* Parent\_Ed} + \alpha\beta_{s2}\text{* Parent\_Ed} + \alpha\gamma\beta_{st2}\text{* Parent\_Ed} +$$

$$\mu_l + \varepsilon_{ijklst}$$

Where i indexes subjects, j indexes classes, k indexes teachers, l indexes schools, s indexes instruction, and t indexes condition.

Significant fixed effects in the final model were:

- Instruction, $F(1, 378.51)= 27.55$, $p<.01$.
- Condition, $F(1, 340.24)=3.985$, $p<.05$.
- Instruction X Condition, $F(1, 340.24)=3.985$, $p<.05$.
- School SES X Instruction, $F(1, 376.38)=23.35$, $p<.01$.
- Parents' Level of Education, $F(1, 412.83)=4.20$, $p<.05$.
- Parents' Level of Education X Instruction X Condition, $F(1, 379.52)=4.65$, $p<.05$.

[12] The author acknowledgement David Nichols of SPSS for guidance through this analysis and for providing the technical language used in much of this subsection. The author takes full responsibility for assumptions, interpretations, etc.

Replicating Figure 13 by using this model for estimation of the regression lines and holding school SES constant at the mean produces a similar picture of trends (see Figure 14):

Consistent:

- Estimated baseline means show little difference between DiSC and placebo groups across all levels of parental education
- Estimated baseline means show a slight increase in baseline challenge/skill across levels of parents' education.
- All estimated e-Mission challenge/skill means are higher than baseline challenge/skill means.
- Estimated e-Mission challenge/skill means for the placebo usage students increase as parents' level of education increases.

Changed for this model:

- Estimated placebo student baseline mean estimate is slightly lower than DiSC study estimate.
- Estimated e-Mission means for the DiSC usage students decrease slightly as parents' level of education increase.
- Estimated baseline means increase in baseline challenge/skill across levels of parents' education is slightly higher.
- Estimated e-Mission challenge/skill means for the placebo usage students whose composite parents' level of education is close to 1 is a bit higher for this model (about 0.25 of a challenge/skill unit).
- The model estimates that control and treatment e-Mission challenge/skill intersect near a parents' level of education of 2.75 (2.83=high school, master's combination).
- The difference between estimated treatment and control e-Mission challenge/skill is larger at the high end of the parents' level of education scale.

The important consistencies for the purposes of this study are (a) the relative positions of the regression lines and (b) the disordinal interaction between parents' level of education and condition represented by the two e-Mission regression lines. Students with lower human capital recorded higher levels of challenge/skill when they used the DiSC tool. This trend seemed to reverse as students' levels of human capital increased. The trend seems to pivot at about 3 (3= a [college, college] combination of parents' level of education.

Scaffolding challenge/skill with DiSC.

Parents' level of education is a measure of human capital at the individual student level. To further examine the effect of the individual students' human capital derived from their parents' level of education, the dataset was disaggregated into less and more human capital at parents' education (less < 3; more $\geq= 3$, see Table 14). School-level SES was not included in the analysis.

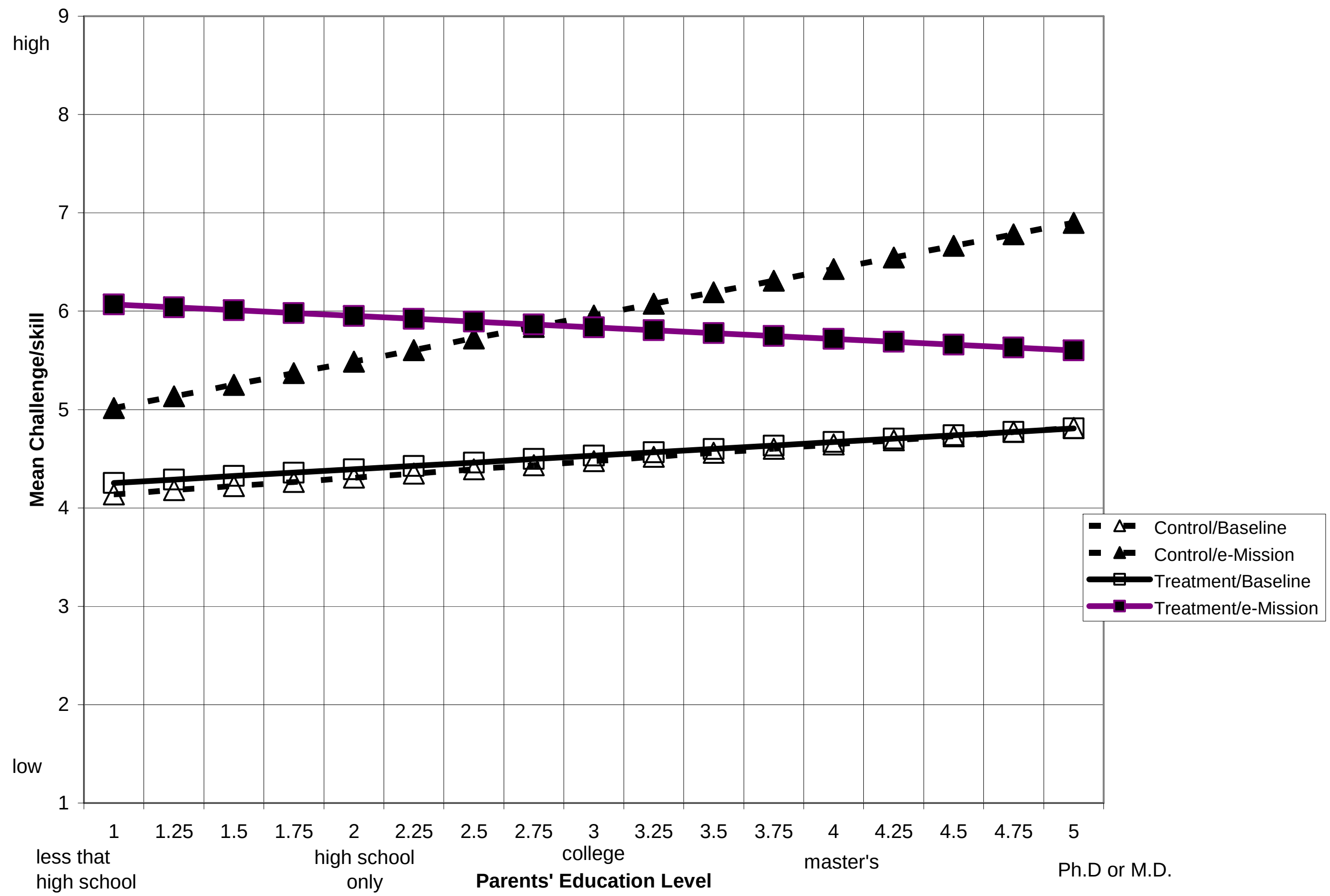


Figure 14. Using Results from a Linear Mixed Models Analysis of Repeated Measures to Estimate Four Regression Lines for Mean Perception of Challenge/Skill Against Parents' Level of Education, One for Each Cell of the Design.
Note: Scale translations for composite parents' level of education (pairs are not ordered): 1=(< high school ,< high school), 1.41=(< high school, high school), 1.73=(< high school, college), 2=(high school, high school) or (< high school, master's), 2.24=(< high school, Ph.D./M.D.), 2.45=(high school, college), 2.83=(high school, master's), 3=(college, college), 3.16=(high school, Ph.D./M.D.), 3.46=(college, master's), 3.87=(college, Ph.D./M.D.), 4=(master's, master's), 4.47=(master's, Ph.D./M.D.), 5=(Ph.D./M.D., Ph.D./M.D.). Composite calculations (geometric mean of mother and father's level of education) drop cases with any missing data.

Figure 15 displays the 95 percent confidence intervals for the e-Mission challenge/skill means for the DiSC and placebo tool usage groups with less and more human capital. The mean difference between DiSC and placebo tool users' challenge/skill for the students with *more* human capital was not significant (see Table 14). The difference between the conditions for students with *less* human capital was significant with a modest effect size, $t(260)=2.94$, $p < .01$, $d = .37$. These results suggest that the DiSC tool benefited students with less human capital. Although the interaction and regression line estimates in the two repeated measures analyses above also indicate that students with more individual human capital achieved higher e-Mission challenge/skill when they used the placebo tool than did students who used the DiSC tool, this result was not significant when this group of students was disaggregated from the larger tool usage sample.

Together, the results of the repeated measures ANOVA, the repeated measure growth model using the mixed model analysis, and the t-test results confirm and elaborate the picture suggested above within the scatterplots for students with low human capital (see Figure 12). The distribution of flow graphed at the high end of human capital by condition was not a significant effect. The distribution at the lower end of human capital was significant.

Table 14.
Descriptives and *t*-test for E-Mission Challenge/Skill by Condition[1] for Students with Less and More Human Capital[2]

Human Capital	*n*	$\overline{X}$	*SD*	*T(df)*	$\overline{X}_{difference}$	Cohen's *d*
Less						
Control	127	5.24	2.34	—	—	—
Treatment	135	6.06	2.16	2.94(260)**	0.82	.37
More						
Control	114	6.04	2	—	—	—
Treatment	90	6.21	2.1	-0.6(202)	-0.17	—

** = $p < .01$
[1]Control = placebo tool, treatment = DiSC tool usage. [2] Human capital is measured by parents' level of education. More human capital = (parents' level of education ≥ 3); Less human capital = (parents' level of education < 3).

A Few Qualitative Comments from Participating Teacher Research Partners

DiSC 2005 was a quantitative study, but the COTF did archive comments from participating educators. A few are included here. All comments were unsolicited. These comments evidence the willingness and commitment of teachers who implemented the study with high fidelity. They engaged as invested Inspiration Challenge research partners.

- Comment 1.
  Hello again,

  First, thank you for the lovely letter you sent about my participation in the Montserrat mission and the development of the DiSC tool as well as the plaque for getting my information to you. I enjoyed the whole process and certainly am thrilled that I got to share this experience with my class. I particularly hope that at some point I could have a copy of the tape that had the discussion tool. I am sure that my students learned a great deal about how to discuss and work as a group from the tape. It certainly was one of the best tools for explaining how a real working discussion should be held that I have ever come in contact with. If I had a copy, I could use it as a refresher after showing it the first time.

I would also like to be added to your list to continue assisting you in any way I can for continuing development of the tool and any future assistance you may need with developing exciting and useful tools for the classroom.

Finally, students from my [my class and my colleagues] went to a school board meeting and made a presentation about the Mission and how much they learned and enjoyed the opportunity. They did a great job and hopefully, this will spur the board to find money so more students can have the same experience. I am hoping that I can provide this mission again for next year's students.

Again, thank you for this great opportunity, and I look forward to hearing from you again.

- Comment 2.
  It truly was the best project I ever did with kids, and I look forward to more opportunities like this for my students. It was an experience they will never forget, and they are ALL so very proud of themselves.

- Comment 3.
  THANK YOU for the wonderful experience. Our students are bouncing off the walls they are so excited and motivated.

  It was a great opportunity for me as well. If interested, I have a few troubleshooting-type thoughts to make the program more teacher-friendly and therefore more marketable.

  Take care and enjoy the holiday season.

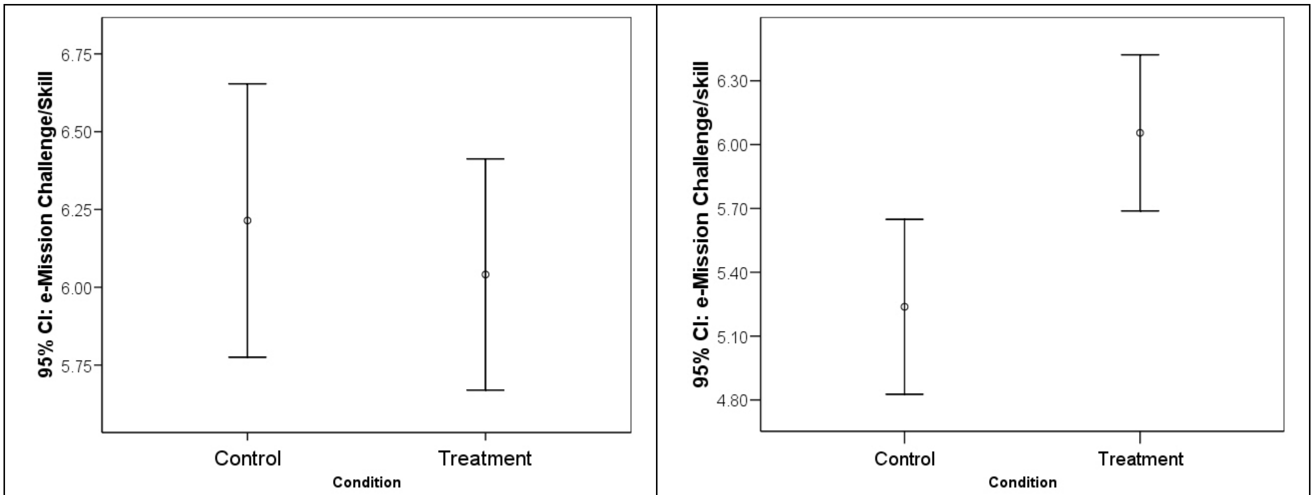
- Comment 4 (excerpts from a longer narrative).

  As a teacher in an inner-city school with many students who are struggling academically, financially, socially, and behaviorally this mission was a great opportunity to get students involved with hands-on real life science. It was with excitement that I pursued this venture….

  Often students would come to me during their study hall period, lunch time, or after school to continue researching [for the Operation Montserrat unit]….

  This experience transformed my mindset to how I teach science to my students….

  Overall, I give this project a 10 out of 10. It has motivated me to continue to seek out learning experiences that challenge, excite, and give students confidence to work hard in school and look to their futures. Students definitely gained a greater interest in science, mathematics, and technology and clearly saw how all the subjects depended on one another….



(a) High human capital | (b) Low human capital

Figure 15. The e-Mission Challenge/Skill[a] Mean score and 95% Confidence Intervals by Condition[b] and Degree of Human Capital[c]
Note: [a]Challenge/Skill = sqrt (Challenge X skill) of ESM responses aggregated at the student level by period. Judgments were made on a 9-point scale (1= *low* , 9 = *high*). [b]Control = placebo tool usage, Treatment = DiSC tool usage. Human capital is measured by parents' level of education. More human capital: parents' level of education ≥= 3; Less human capital: parents' level of education < 3.

Conclusions

The DiSC 2005 Inspiration Challenge results provide a baseline to which COTF can compare subsequent iterations of the DiSC tool in practice of both formative evaluation and revision and design research. These analyses concentrated on changes because of DiSC, changes over time due to instruction, change because of the e-Mission live simulation, and the covariate of parent's level of education. Many of the findings were derived through a variety of complementary methods (e.g., graphical and statistical) including a model that nests students within their educational context (classroom, teacher, and school). During the statement of conclusions, the challenge/skill measures are reframed based upon the assumptions within the Methodology section that argued e-Mission challenge/skill measures e-Mission states and dimensions of experience at time = 55 minutes. Findings discussed within this section will be stated using the construct "flow" or one of the other three states of experience replacing the challenge/skill label:

1. Dummy coding from *z* score
   - *Flow*: High challenge/High skill = (+ challenge/ +skill).
   - *Anxiety*: High challenge/Low skill = (+challenge - skill).
   - *Relaxation*: Low challenge/High skill = (-challenge/+ skill).
   - *Apathy*: Low challenge/Low skill = (-challenge/- skill).

2. An individual's geometric mean of skill and challenge indicators
   - *Flow:* Challenge/Skill $= (challenge * skill)^{1/2}$.

DiSC 2005 was designed to address three hypotheses:

Hypothesis 1. DiSC usage will increase student achievement along flow and its dimensions: mental models and self-efficacy.

DiSC usage increased flow. There is a direct relationship between flow and student achievement. Students with higher flow scored better on the standards-based test.

During this iteration DiSC did not enhance academic self-efficacy, social self-efficacy, or argumentation self-efficacy. Scores for students who used the DiSC tool decreased for nature of science and job knowledge; scores for students who used the placebo tool did not decrease.

Hypothesis 2. There is an interaction between DiSC tool usage and human capital in predicting flow.

Parents' level of education is a student-level measure of human capital. There was an inverse relationship between human capital and flow for the DiSC students. The lower parents' level of education, the more DiSC scaffolded learners' flow. DiSC appears to have scaffolded students with less human capital. When the student sample was disaggregated into two groups by parents' level of education (more and less education), there was an inverse relationship between human capital and flow for the subsample of students who reported their parents had less education. For this group and on the average, the less human capital a student had reported, the more DiSC usage increased that student's e-Mission flow.

Hypothesis 3. NASA-approved product e-Mission Operation Montserrat will increase student achievement along flow and its dimensions: mental models and self-efficacy.

COTF has long assumed that the e-Mission live simulation presents a qualitatively and quantitatively more engaging and challenging experience than typical classroom activity. The strongest DiSC 2005 effect was the finding that the e-Mission experience increases flow. This result supports the position that the live simulation is a substantively different learning environment (from the perspective of flow, apathy, relaxation, and anxiety) than classroom conditions that existed at baseline and waves 2-4. This finding was accompanied by the weaker effect for the DiSC tool: Students who had used the DiSC tool experienced higher e-Mission flow than students who had used the placebo tool. Increased flow during the e-Mission was accompanied by decreased frequency in reports of apathy (low skills and challenges). All students made significant gains from pretest to posttest on the curriculum-oriented exam.

Discussion

Bell (2004) wrote that over the course iterations implementing the SenseMaker tool and its predecessors, he and his colleagues "never dramatically improved the number of students that developed an integrated understanding about the specific debate topic" (p. 123). However, by iteration six Bell and his colleagues were deriving "compelling" results for gains in understanding because of modifications engendered by design principles derived from analysis of iterative implementation cycles. One goal of this DiSC iteration is to begin to develop *design principles* that will guide tool development and implementation. Another is *refinement* of the study implementation. A third is to provide formative evaluation for *revision* of the instructional unit.

Results suggest:

- Although they engaged in the roles of Earth system scientists, students did not build a knowledge base of the training required for and characteristics of Operation Montserrat-related STEM careers. It could be that job knowledge was not sufficiently covered during the instruction. It could be that it was understated within the curriculum or that participating teachers did not emphasize job knowledge. In any case, it appears students need more explicit exposure than they had during DiSC 2005. COTF's Challenger Learning Center has since developed a job knowledge component that can be incorporated into instructional implementation for DiSC 2006. Participating teachers will need to be aware that job knowledge is an instructional goal for the unit.

Implementation refinement: Alert teachers to study goals and the curricular components that support them and ensure alignment between the assessment instrument and the instruction.
Unit revision: Incorporate job knowledge instruction into unit.
Design principle candidate: Reinforce Operation Montserrat activities through classroom discussion that makes unit components explicit and includes components within classroom practice.

- If the nature of science scale ran from 0 to 100 percent, these students scored about 60 percent on both pre- and postmeasures across both conditions. The increase in flow during the e-Mission supports the conclusion that DiSC enhanced students' ability to engaged in the practice of science as they applied the knowledge and skills they gained during the Operation Montserrat unit in new and challenging ways. Overall gains in flow due to the e-Mission were even higher than gains due to tool usage. However, engagement in practice of science and argumentation did not seem to transfer over to the paper and pencil items that measured student argumentation self-efficacy or their mental model of the nature of scientific enterprise. Perhaps paper and pencil measures are not optimal measures for this type of learning (Lederman et al., 2002). However, it is also possible that students lacked the self-reflective and analytical (metacognitive) ability to recognize their practice when it was described in prose. The science reform movement stresses that classroom norms should reflect assessment goals within the discourse of the learning community (Hernandez-Gantes & Reese, 2004). The DiSC argumentation rubric (Reese & McFarland, 2006) was designed to build students' metacognitive awareness of scientific discourse practice. It is possible that DiSC 2005 teachers did not make the most of this component of the tool.

Implementation refinement: Alert teachers to study goals, the curricular components that support them, and how to infuse both within classroom practice as a learning community norm.
Design principle candidate: Scaffold learners' metacognitive awareness of their engagement in scientific practice through tools (like the DiSC argumentation rubric) and teacher-student and student-student classroom discourse.

- Traditional instructional design assumes a manifesto that properly designed instruction will result in learner gains at targeted learning goals and objectives when design, development, and assessment are aligned and targeted toward learner characteristics (Gagné, Briggs, & Wager, 1992; Smith & Ragan, 1993). More recent perspectives in educational psychology suggest that learning environments should be situated in social contexts that allow individuals to engage as learning communities to co-engage in practices and frames of mind authentic to how targeted learning goals are engaged in authentic enterprise (Greeno et al., 1996; Schoenfeld, 2006). Both perspectives "support structures" (Schoenfeld, 2006, p. 17) like the DiSC technology tool that "are helpful in various implementation contexts," such as those of the underachieving students who attend the NASA Explorer Schools. DiSC 2005 flow results indicate that DiSC scaffolded skills and ability to engage in challenge for students with low human capital, in effect, leveling the playing field for students across their parents' education levels. This effect should be studied. In addition to teasing out how DiSC scaffolds e-Mission flow, future iterations should identify (a) what obstacles DiSC presented to students with more human capital and (b) how to overcome those obstacles.

Implementation refinement: Use a study design that allows researchers a finer lens to study learner characteristics and their engagement with the DiSC tool. While later iterations should replicate DiSC 2005 with large-scale implementation at a distance, these results motivate a DiSC 2006 design that affords classroom observation during DiSC utilization.

- Curriculum implementation and the research that studies it depends on the "character of the implementation in context" (Schoenfeld, 2006, p. 17). When educators serve as research partners, they must be "trained" and "invested" in both the curriculum and study protocols. The unsolicited communications sampled within the brief section of qualitative comments demonstrates the enthusiasm and dedication of DiSC 2005 Inspiration Challenge research partners who implemented the study with the fidelity required to be included within the tool usage dataset employed for most of the study analyses. Seventy teachers had volunteered to participate in the DiSC 2005 Inspiration Challenge; 50 were randomly selected. Forty-one teachers representing 50 classrooms received study materials. Of these, 36 teachers representing 45 classrooms actually participated. However, only 26 teachers representing 30 classrooms could be included within the tool usage dataset. This is an unfortunate attrition rate, especially when costs of COTF staff facilitation and administration time and shipping and supply are considered. Given (a) the time constraints of DiSC 2005 (the length of the contract and the late start because of the funding disbursement date), (b) the requirement to target participants from the NASA Explorer Schools testbed, and (c) the advisement by consultants that analyses would require 30-50 teachers to ensure sufficient power, perhaps the attrition was unavoidable. As things turned out, the 13 control and 17 treatment teachers with adequate implementation fidelity provided enough power to tease out the small effect sizes that would be expected to accompany a first iteration design study.

Implementation refinement: Future contract, funding, and implementation cycles should be designed to support study success. The DiSC 2005 Inspiration Challenge timeline was much too short, and it remains a testament to the quality, talent, and dedication of the COTF staff that they designed, developed, implemented, and analyzed tools and studies that resulted in this rich array of findings, goodwill, and possibilities for the future.

Limitations

The fact that the e-Mission experience (at time = 55) inspired such gains in the states of experience (increased flow, decreased apathy, decreased relaxation, and increased anxiety) actually jeopardized the integrity of the study design because the e-Mission ESM had to be analyzed as its own period rather than as a component of wave 4. The assumptions underlying interpretation of the e-Mission ESM data must be accepted to apply an interpretation involving the states and qualities of experience. DiSC 2006 will investigate the design of a shortened ESM form that might be administered more often during the e-Mission to collect a larger number of responses for that period.

Remaining DiSC 2005 Opportunity

Results are very clear that students' state of experience during the e-Mission (at least at time = 55 minutes) was quantitatively different from the average experiences reported over the other four periods (up to 29 ESM reports). ESM data for all of the quality of experience dimensions (e.g., happiness, enjoyment, self-esteem, and salience) remain to be analyzed, as do the survey 1 data describing students' overall affective and cognitive traits and context parameters (such as the parents' level of education, which we analyzed within this report). That remaining data had the potential to inform the findings presented within this report. .

Synthesis

As the Model of Systemic Inspiration Growth derived from the synthesis of diverse research literatures, a concept of inspiration emerged in which learning environments might support young people in making the types of productive life choices that lead to science, technology, engineering, and mathematics (STEM) literacy and career pipeline by focusing on dimensions known to increase intrinsic satisfaction when individuals surmounted learning and life challenges through increased procedural and declarative knowledge. One salient point that emerged during 2005 COTF interviews of NASA education leaders for an expert panel needs assessment report (Reese, 2005) was that viable educational research must be conducted in authentic classrooms that mirror the problems facing students and teachers in today's U.S. classrooms. By design the baseline iteration of DiSC inspiration research was conducted in schools with high proportions of low achieving, underserved student populations. By design the Inspiration Challenge was implemented in schools with identified academic, technological, and sociocultural challenges. Alan H. Schoenfeld, president of the American Educational Research Association in 1999, advised (2006) that research must provide differentiated results. It must:

> Provide evidence of what the curriculum's effects might be in varied contexts—and, more important, evidence of what kinds of support structures are helpful in various implementation contexts (i.e., urban, rural, or suburban schools, schools with high proportions of second language learns, etc. (p. 17).

The DiSC tool, e-Mission Operation Montserrat, and DiSC 2005 concern a unit of instruction, a learning context/delivery system, and a tool that provides scaffolding. The human capital results provide evidence for what effect DiSC might have in contexts of low individual human capital. The e-Mission flow results provide evidence of the effects of live simulations delivered at a distance via the Internet. These significant baseline results suggest that the DiSC tool holds promise for increasing science academic achievement and inspiration for this population of students. Results suggested that DiSC 2005 helped to enfranchise participating NASA Explorer Schools students with lower amounts of human capital.

References

Bandura, A. (2001). Guide for constructing self-efficacy scales (Revised). Retrieved September 7, 2004, from Available from Frank Pajares, Emory University.

Bell, P. (1997, December). *Using argumentation representations to make thinking visible.* Paper presented at the Computer-supported Collaborative Learning Conference, Toronto.

Bell, P. (1998). *Designing for students' science learning using argumentation and classroom debate.* Unpublished dissertation, University of California, Berkeley.

Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M.C. Linn, E.A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 115-143). Mahweh, NJ: Lawrence Erlbaum Associates.

Coleman, J.S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology, 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure*, S95-S120.

Csikszentmihali, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *The Journal of Nervous and Mental Disease, 175*(9), 526-536.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (HarperPerennial ed.). New York: Harper & Row Publishers Inc.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention.* New York: HarperCollins Publishers.

Csikszentmihalyi, M. (1997). Flow and creativity. *NAMTA Journal, 22*(2), 60-97.

Csikszentmihalyi, M., & Schneider, B. (2000). *Becoming adult: How teenagers prepare for the world of work* (1st ed.). New York, NY: Basic Books.

Csikszentmihalyi, M., & Schneider, B. (2001). Conditions for optimal development in adolescence: An experiential approach. *Applied Developmental Science, 5*(3), 122-124.

Design-based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Gagné, R.M., Briggs, L.J., & Wager, W.W. (1992). *Principles of instructional design.* New York: Harcourt Brace Jovanovich College Publishers.

Gee, J.P. (2001). Identity as an analytic lens for research in education. *Review of Research in Education, 25*, 99-125.

Gilbert, J.K., & Boulter, C.J. (Eds.). (2000). *Developing models in science education.* Boston: Kluwer Academic Publishers.

Greeno, J.G., Collins, A., & Resnick, L.B. (1996). Cognition and learning. In D.C. Berlinger & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York: Macmillan.

Hernandez-Gantes, V., McGee, S., Reese, D.D., Kirby, J., & Martin, J. (2004). *NASA Explorer Schools evaluation brief 3: A program in the making.* Wheeling, WV: NASA-sponsored Classroom of the Future, Erma Ora Byrd Center for Educational Technologies, Wheeling Jesuit University.

Hernandez-Gantes, V., & Reese, D.D. (2004). *Classroom observation guidelines: InSTEP™ (Integrating Strategies and Technology in Educational Practice).* Wheeling, WV: Wheeling Jesuit University.

Hickey, D.T., Kindfield, A.C.H., Horwitz, P., & Christie, M.A.T. (2003). Integrating curriculum, instruction, assessment, and evaluation in a technology-supported genetics learning environment. *American Educational Research Journal, 40*(2), 495-538.

Hickey, D.T., Recesso, A., Hay, K., Zuiker, S., Wallace, C., & Peker, D. (2003, April). *We know it when we see it: Video-supported formative assessment of inquiry-oriented activity and instruction.* Paper presented at the GSTEP Research Group, University of Georgia, Athens.

Hickey, D.T., Zuiker, S.J., & McGee, S. (2004, April). *A multi-level/multi-type model for design-based balancing of formative and summative assessment.* Paper presented at the American Educational Research Association, San Diego.
Howard, B. (2004). *Results from phase two research on e-Mission pre- and posttests*. Wheeling, WV: Wheeling Jesuit University.
Kelly (Ed.), A.E. (2003). Theme issue: The role of design in educational research. *Educational Researcher, 32*(1).
Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education 77*(3), 319-337.
Lederman, N.G., Abd-El-Khalick, F., Bell, R.L., & Schwartz, R.S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497-521.
Lipsey, M.W. (1990). *Design sensitivity*. Newbury Park, CA: SAGE Publications.
Lobato, J. (2003). How design experiments can inform a rethinking of transfer and vice versa. *Educational Researcher, 32*(1), 17-20.
Luke, D.A. (2004). *Multilevel modeling*. Thousands Oaks, CA: A Sage University Paper.
National Aeronautics and Space Administration Office of Education. (2003). *Education Enterprise strategy*. Washington, DC: NASA Office of Education.
National Research Council. (1995). *National science education standards*. Washington, DC: National Academy Press.
National Research Council. (1999). *The assessment of science meets the science of assessment: Summary of a workshop*. Washington, DC: National Academy Press.
National Research Council. (2001). *Knowing what students know*. Washington, DC: National Academy Press.
Pedhazur, E.J. (1997). *Multiple regression in behavioral research*. New York: Harcourt Brace College Publishers.
Portes, A., & MacLeod, D. (1996). Educational progress of children of immigrants: The roles of class, ethnicity, and school context. *Sociology of Education, 69*, 255-275.
Reese, D.D. (2003). *Metaphor and content: An embodied paradigm for learning*: Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University.
Reese, D.D. (2005). *Needs assessment phase II: NASA expert panel report (responses and recommendations)* (report No. NA/R2/Dec-2005). Wheeling, WV: Center for Educational Technologies, Wheeling Jesuit University.
Reese, D.D., Kim, B., Palak, D., Smith, J., & Howard, B. (2005). *Inspiration brief 1: Defining inspiration, the Inspiration Challenge, and the informal event (concept paper)* (No. COTF/IB1/6-2005). Center for Educational Technologies, Wheeling, WV: Wheeling Jesuit University.
Reese, D.D., & McFarland, L. (2006). *Inspiration brief 2: The DiSC and RoboKids tools and labs (design and testing)* (No. COTF/B2/Jan-2006). Wheeling, WV: Center for Educational Technologies, Wheeling Jesuit University.
Reese, D.D., Smith, J., & McFarland, L. (2005). e-Mission Montserrat, part 1: Achievement findings from 2004. Wheeling, WV: Challenger Learning Center, Center for Educational Technologies, Wheeling Jesuit University.
Ruiz-Primo, M.A., Shavelson, R.J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5), 369-393.
Schafer, N.J., Kruger, A.C., & Hickey, D.T. (n.d.). Video feedback to facilitate argumentation. University of Georgia, Athens.
Schneider, B. (1993). Year 5 Sloan study.

Schoenfeld, A.H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35*(2), 3-22.

Short, J.F. (1997). *Poverty, ethnicity, and violent crime*: Westview Press: A Division of HarperCollins Publishers Inc.

Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24*(4), 323-355.

Singley, M.K., & Anderson, J.R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Smith, P.L., & Ragan, T.J. (1993). *Instructional design* (1st ed.). New York: Merrill, an imprint of Macmillan Publishing Company.

SPSS. (2005). Statistics Coach (Vol. 2005). SPSS.

Taasoobshirazi, G., & Hickey, D.T. (2005). *Design-based implementation and evaluation: Astronomy Village: Investigating the Universe* (final implementation report).

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research articles: A simplified methodology. Retrieved March 30, 2006, from http://work-learning.com/effect_sizes.htm

UCLA Academic Technology Services. (2006). SPSS paper examples: Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth models, by Judith Singer. *Downloadable Papers and Books on Statistical Computing.*

Appendix A: School Demographics

The table lists the school demographics for all classes that received Inspiration Challenge study materials. Some schools housed multiple Inspiration Challenge classes.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials

School Demographics[1]	Class A	Class B	Class C[6]	Class D[6]	Class E
Number of Students	1262	1262	305	305	509
State	NV	NV	OR	OR	HI
NASA Field Center[5]	ARC	ARC	ARC	ARC	ARC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Urban	Urban	Rural	Rural	Rural
School Grades	6-8	6-8	4-8	4-8	6-8
SES%[3]	85%	85%	58%	58%	50%
Title I	yes	yes	no	no	yes
Number of Migrant Students	0	0	16	16	8
Minority%	86%	86%	18%	18%	72%
Black	24%	24%	1%	1%	0%
Hispanic	59%	59%	9%	9%	6%
Asian	3%	3%	0%	0%	65%
American Indian/Alaskan	0%	0%	9%	9%	1%
White	14%	14%	82%	82%	28%
Study Attrition Status[4]	Retained	Retained	Retained	Retained	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] ARC = Ames Research Center. [6] School consolidated in the year 2004 and is now a K-8 school, and, therefore, the related data is obsolete and will need to be updated.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

Table 1

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class F	Class G	Class H	Class I	Class J
Number of Students	509	588	417	699	1404
State	HI	AZ	CA	AZ	NV
NASA Field Center[5]	ARC	DFRC	DFRC	DFRC	ARC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Rural	Urban	Suburban	Rural	Urban
School Grades	6-8	KG-6	6-8	7-8	6-8
SES%[3]	50%	N/A	20%[7]	N/A	86%
Title I	yes	no	no	yes	yes
Number of Migrant Students	8	0	0	0	0
Minority%	72%	24%	35%	39%	84%
Black	0%	1%	17%	2%	22%
Hispanic	6%	13%	9%	16%	59%
Asian	65%	3%	8%	1%	3%
American Indian/Alaskan	1%	7%	1%	20%	1%
White	28%	76%	65%	61%	16%
Study Attrition Status[4]	Retained	Retained	Retained	Retained	Dropped

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] ARC = Ames Research Center, DFRC = Dryden Flight Research Center. [7] 2003-2003 school year data obtained from the school. NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class K	Class L	Class M	Class N	Class O
Number of Students	699	1262	950	950	1262
State	AZ	NV	CA	CA	NV
NASA Field Center[5]	DFRC	ARC	DFRC	DFRC	ARC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Rural	Urban	Suburban	Suburban	Urban
School Grades	7-8	6-8	6-8	6-8	6-8
SES%[3]	NA	85%	66%	66%	85%
Title I	yes	Yes	yes	yes	yes
Number of Migrant Students	0	0	65	65	0
Minority%	39%	86%	66%	66%	86%
Black	2%	24%	25%	25%	24%
Hispanic	16%	59%	38%	38%	59%
Asian	1%	3%	3%	3%	3%
American Indian/Alaskan	20%	0%	0%	0%	0%
White	61%	14%	34%	34%	14%
Study Attrition Status[4]	Retained	Retained	Retained	Retained	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] ARC = Ames Research Center, DFRC = Dryden Flight Research Center.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class P	Class Q	Class R	Class S	Class T
Number of Students	1262	720	647	647	647
State	NV	NY	MA	MA	MA
NASA Field Center[5]	ARC	GSFC	GSFC	GSFC	GSFC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Urban	Urban	Urban	Urban	Urban
School Grades	6-8	6-8	6-8	6-8	6-8
SES%[3]	85%	55%	NA	NA	NA
Title I	yes	yes	yes	yes	yes
Number of Migrant Students	0	N/A	20	20	20
Minority%	86%	45%	25%	25%	25%
Black	24%	32%	9%	9%	9%
Hispanic	59%	8%	11%	11%	11%
Asian	3%	5%	5%	5%	5%
American Indian/Alaskan	0%	0%	0%	0%	0%
White	14%	55%	75%	75%	75%
Study Attrition Status[4]	Retained	Retained	Retained	Dropped	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] ARC = Ames Research Center, GSFC = Goddard Space Flight Center.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class U	Class V	Class W	Class X	Class Y
Number of Students	332	786	496	97	786
State	VT	TX	CA	ND	TX
NASA Field Center[5]	GSFC	JSC	JPL	JSC	JSC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Isolated Rural	Urban	Suburban	Isolated Rural	Urban
School Grades	7-8	6-8	K-6	KG-6	6-8
SES%[3]	55%	75%	47%	66%	75%
Title I	yes	yes	no	yes	yes
Number of Migrant Students	15	3	0	0	3
Minority%	5%	70%	83%	100%	70%
Black	2%	17%	13%	0%	17%
Hispanic	1%	52%	65%	0%	52%
Asian	2%	1%	4%	0%	1%
American Indian/Alaskan	1%	0%	1%	100%	0%
White	95%	30%	17%	0%	30%
Study Attrition Status[4]	Retained	Retained	Retained	Dropped	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] GSFC = Goddard Space Flight Center, JSC = Johnson Space Center, JPL = Jet Propulsion Laboratory.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class Z[2]	Class AA	Class AB	Class AC	Class AD
Number of Students	450	818	818	823	1262
State	NM	NM	NM	GA	NV
NASA Field Center[5]	JSC	JSC	JSC	KSC	ARC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Isolated Rural	Isolated Rural	Isolated Rural	Urban-Inner City	Urban
School Grades	K-12	6-8	6-8	6-8	6-8
SES%[3]	100%	58%	58%	84%	85%
Title I	yes	yes	yes	yes	yes
Number of Migrant Students	NA	11	11	0	0
Minority%	100%	67%	67%	100%	86%
Black	0%	2%	2%	99%	24%
Hispanic	0%	64%	64%	0%	59%
Asian	0%	0%	0%	0%	3%
American Indian/Alaskan	100%	1%	1%	0%	0%
White	0%	33%	33%	0%	14%
Study Attrition Status[4]	Retained	Dropped	Dropped	Retained	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable neither in part or overall data analysis or did not return any parental consent forms. [5] ARC = Ames Research Center, JSC = Johnson Space Center, KSC = Kennedy Space Center.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, may not add up to 100 percent ethnicity.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class AE	Class AF	Class AG	Class AH	Class AI
Number of Students	823	713	746	327	568
State	GA	FL	SC	KY	SC
NASA Field Center[5]	KSC	KSC	LRC	LRC	LRC
School Type	Public	Public (magnet)[2]	Public	Public	Public
School Locale[2]	Urban-Inner City	Urban	Rural	Rural	Rural
School Grades	6-8	6-8	6-8	6-8	6-8
SES%[3]	84%	61%	70%	99%	46%
Title I	yes	Yes	yes	yes	no
Number of Migrant Students	0	0	0	40	1
Minority%	100%	68%	50%	1%	41%
Black	99%	31%	49%	0%	38%
Hispanic	0%	33%	1%	0%	3%
Asian	0%	4%	0%	0%	0%
American Indian/Alaskan	0%	0%	0%	0%	0%
White	0%	32%	50%	99%	59%
Study Attrition Status[4]	Retained	Retained	Retained	Retained	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] KSC = Kennedy Space Center, LRC = Langley Research Center.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class AJ	Class AK	Class AL	Class AM	Class AN
Number of Students	717	546	487	417	219
State	SC	KY	AZ	CA	SC
NASA Field Center[5]	LRC	LRC	DFRC	DFRC	LRC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Rural	Suburban	Isolated Rural	Suburban	Rural
School Grades	6-8	6-8	PK-8	6-8	7-12
SES%[3]	73%	99%	NA	20%[7]	59%
Title I	yes	Yes	yes	no	no
Number of Migrant Students	11	11	0	0	0
Minority%	69%	47%	17%	35%	33%
Black	63%	34%	1%	17%	32%
Hispanic	5%	8%	12%	9%	2%
Asian	0%	5%	1%	8%	0%
American Indian/Alaskan	0%	1%	3%	1%	0%
White	31%	53%	83%	65%	67%
Study Attrition Status[4]	Dropped	Retained	Retained	Retained	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] DFRC = Dryden Flight Research Center, LRC = Langley Research Center. [7] 2003-2003 school year data obtained from the school. NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

School Demographics for All Classes Receiving Inspiration Challenge Study Materials (cont.)

School Demographics[1]	Class AO	Class AP	Class AQ	Class AR	Class AS
Number of Students	331	1262	1262	826	826
State	IA	NV	NV	MS	MS
NASA Field Center[5]	MSFC	ARC	ARC	SSC	SSC
School Type	Public	Public	Public	Public	Public
School Locale[2]	Isolated Rural	Urban	Urban	Suburban	Suburban
School Grades	6-12	6-8	6-8	4-8	4-8
SES%[3]	27%	85%	85%	35%	35%
Title I	no	yes	yes	no	no
Number of Migrant Students	0	0	0	0	0
Minority%	3%	86%	86%	17%	17%
Black	1%	24%	24%	16%	16%
Hispanic	2%	59%	59%	1%	1%
Asian	1%	3%	3%	0%	0%
American Indian/Alaskan	0%	0%	0%	0%	0%
White	97%	14%	14%	83%	83%
Study Attrition Status[4]	Retained	Retained	Retained	Retained	Retained

[1] Data obtained through the National Center for Education Statistics (NCES) http://nces.ed.gov/ccd/schoolsearch , Common Core of Data (CCD) public school data for the 2002-2003 school years unless otherwise specified. [2] Data obtained through the Center for Educational Technologies (CET). [3] Socioeconomic status percentage derived from NCES number of free lunch eligible plus reduced-price lunch eligible divided by total students unless otherwise specified. [4] Retained = Data used in analysis, Dropped = Participants who did not complete the study with enough data to make them usable either in part or overall data analysis or did not return any parental consent forms. [5] ARC = Ames Research Center, MSFC = Marshall Space Flight Center, SSC = Stennis Space Center.
NOTE: All percentages have been rounded to the nearest whole number, and, therefore, ethnicity may not add up to 100 percent.

Appendix B: Crosstabs of Student-reported Mother's Level of Education by Father's Level of Education

Mother's Education		Father's Education					Total
		Less Than High School	High School Only	College	Master's	Ph.D. or M.D.	
Less Than High School	Count	40	20	9	0	0	69
	% within mother's education	58	29	13	0	0	100
	% within father's education	67	9	5	0	0	12
	% of total	7	4	2	0	0	12
High School Only	Count	12	126	35	3	3	179
	% within mother's education	7	70	20	2	2	100
	% within father's education	20	56	18	5	13	32
	% of total	2	22	6	1	1	32
College	Count	6	56	131	19	4	216
	% within mother's education	3	26	61	9	2	100
	% within father's education	10	25	68	31	17	38
	% of total	1	10	23	3	1	38
Master's	Count	1	17	13	32	5	68
	% within mother's education	2	25	19	47	7	100
	% within father's education	2	8	7	53	22	12
	% of total	0	3	2	6	1	12
Ph.D. or M.D.	Count	1	5	6	7	11	30
	% within mother's education	3	17	20	23	37	100
	% within father's education	2	2	3	12	48	5
	% of total	0	1	1	1	2	5
Total	Count	60	224	194	61	23	562
	% within mother's education	11	40	35	11	4	100
	% within father's education	100	100	100	100	100	100
	% of total	11	40	35	11	4	100

Appendix C: ESM Administration Directions

This appendix contains the direction sheets for ESM #1 and ESM #30 (the e-Mission ESM).

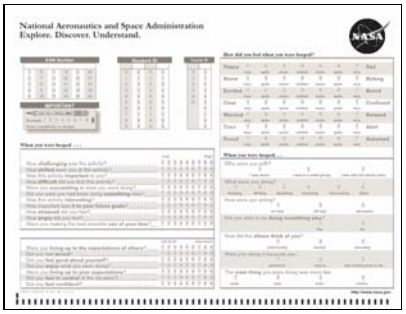


Figure 1. ESM Instrument

ESM #1

Set the timer for 7 minutes.

Use these directions to administer ESM #1

1. You should have assigned each student to a student ID number on the ID sheet provided in your mailing. These numbers are very important, and each student must use the same number for everything they submit for the Inspiration Challenge. The mailing also contained a sheet of labels printed with those student ID numbers. Distribute labels containing the Student ID numbers to the correct students. You should give each student the label with his/her ID number to affix to a card, paper, etc. You should have your students stick their labels to a paper, a card, or folder, etc.. Tell the students how and where you want them to affix their labels. If a student misplaces an ID, use your ID list to give the student another copy of the same number.

2. Say: **This is your Inspiration Student ID number. You will keep and use the same number throughout the Inspiration study. If you forget or lose your number, tell me and I will give you another copy of your number. You must use the same number for the entire Inspiration Challenge.**

3. Say: **Each day of the Inspiration study, when the timer rings, it will be time for the ESM.**

4. When the timer rings, say: **It is time for the Inspiration ESM. You will use the ESM to record exactly how you felt and thought when the timer rang.**

5. Distribute one ESM sheet to each student and say: **This is the ESM form. You will write your answers on this form by filling in the rectangles.**

6. Say: **You must use a #2 pencil to write on the ESM form. Do not use a pen.** Give pencils to students who do not have pencils.

7. Write "ESM #1" on the board.

8. Say: **"Find the rectangle labeled "ESM Number" and write "1" on your ESM Scantron sheet below where it says "ESM Number." Fill in the little rectangle numbered "1" in the ESM rectangle below the number you wrote. Make sure you fill in each rectangle completely.**

9. Say: **Find the rectangle labeled "Student ID." Write your Student ID on your ESM sheet where it says Student ID. Put one ID number in each of the four squares in the**

top row. Then fill in the little rectangle in the column beneath each number that is the same as the number you wrote above it. Make sure you fill in each rectangle completely.

10. Write your Teacher ID number on the board. Say: **Find the rectangle labeled "Teacher ID" and write my Teacher ID number on the ESM sheet where it says "Teacher ID." Put one number in each of the two columns. Then fill in the little rectangle in the column beneath each number that is the same as the number you wrote above it. Make sure you fill in each rectangle completely.**

11. Say: **Today we are going to work together to fill out the ESM. For the next month you will fill out one or two ESMs every class. After today you will complete the ESM by yourself. As we work together today, please ask me to explain any words or items you don't understand.**

12. Say: **Look at the questions in the left-hand box under the pencil. The first one asks you how challenging the activity was when you were beeped. If you think it was not challenging because it was too easy, you would fill in the rectangle with the "1." If it was very challenging because it was very hard for you, you would fill in the number "9' under the word "hard.' If it was some place in between the low challenge and hard challenge, fill in the number that tells you how close it is to low challenge or high challenge. If it was right in the middle, you would fill in a "5."**

13. Lead the students through the rest of this box in this manner. If the students seem to understand the items, you can go more quickly.

14. Say: **Look at the questions in the lower left-hand box of the ESM. Find the first ESM item in that box: "Were you living up to the expectations of others?" I'm going to explain this item to you. When other people, like your teachers, friends, and family, want you to act in certain ways, they have expectations for you. When you act the way they think you should, you are living up to the expectations of others. So, when the timer beeped, were you acting and thinking the way others think you should?** Provide more assistance if the students still don't understand this item.

15. Say: **There is another item in this box that asks, "Were you living up to <u>your</u> expectations?" This item asks if you were acting and thinking the way <u>you</u> think you should when the timer beeped.**

16. Lead the students through the rest of the ESM in this manner. If the students seem to understand the items, you can go more quickly.

17. Collect the completed ESM forms and resume classroom activities.

18. When you have time after class, make sure the students have correctly completed the form. Check the ESM number, the Student ID, and the Teacher ID. If there are mistakes, correct them and be sure to review the correct procedure before administering the next ESM.

19. Place this ESM set in the box we provided for mail pickup at your main office on Monday, October 3. On October 3 mail the five ESM sets (numbers 1-5) together with one Pre-Inspiration Survey 1 set, one Pre-Inspiration Survey 2 set, and one Pre-test Curriculum-oriented Exam.

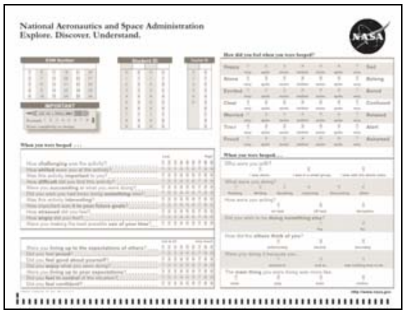

National Aeronautics and Space Administration
Explore. Discover. Understand.

Figure 1. ESM Instrument

ESM #30

Set the timer for <u>**55**</u> minutes.

Use these directions to administer ESM #30 (see Figure 1).

1. When the timer rings, say: **It is time for the Inspiration ESM. You will use the ESM to record exactly how you felt and thought <u>when the timer rang</u>.**
2. Distribute one ESM sheet to each student, and say: **This is the ESM form. You will write your answers on this form by filling in the rectangles.**
3. Say: **You must use a #2 pencil to write on the ESM form. Do not use a pen.** Give pencils to students who do not have pencils.
4. Write **"ESM #30"** on the board.
5. Say: **"Find the rectangle labeled "ESM Number" and write "30" on your ESM Scantron sheet below where it says "ESM Number." Fill in the little rectangle numbered "30" in the ESM rectangle below the number you wrote. Make sure you fill in each rectangle completely.**
6. Say: **Find the rectangle labeled "Student ID." Write your Student ID on your ESM sheet where it says Student ID. Put one ID number in each of the four squares in the top row. Then fill in the little rectangle in the column beneath each number that is the same as the number you wrote above it. Make sure you fill in each rectangle completely.**
7. Write your Teacher ID number on the board. Say: **Find the rectangle labeled "Teacher ID" and write my Teacher ID number on the ESM sheet where it says "Teacher ID." Put one number in each of the two columns. Then fill in the little rectangle in the column beneath each number that is the same as the number you wrote above it. Make sure you fill in each rectangle completely.**
8. Say: **Please use a pencil and complete the ESM.**
9. Collect the completed ESM forms and resume classroom activities.
10. When you have time after class, make sure the students have correctly completed the form. Check the ESM number, the Student ID, and the Teacher ID. If there are mistakes, correct them and be sure to review the correct procedure before administering the next ESM.
11. Place this ESM set in the box we provided for mail pickup at your main office. This is the final ESM administration. Well done! Thank you!!!

Appendix D: The Curriculum-oriented Exam and Standards-based Test

A team of graduate students[13] trained under Dr. Daniel T. Hickey contracted with COTF to prepare the curriculum-oriented exam and standards-based test according to Hickey's multilevel assessment technique. Appendix D contains portions of the materials they developed for DiSC 2005, specific to the Operation Montserrat unit. COTF later modified the quizzes and answer explanations (renamed topic summaries). The curriculum-oriented exam and the standards-based test multiple choice items were used as prepared by the assessment team with the exception that a fifth "I don't know" alternative was added when items provided only four alternatives.

Contents

- Introduction
- Alignment Framework
- Curriculum Outline

[13] University of Georgia Learning and Performance Support Laboratory (LPSL) Operation Montserrat Assessment Team: Dionne I. Cross, Gita Taasoobshirazi, Kate Anderson, Steven J. Zuiker

Introduction

Considering its different roles in science education, a National Research Council (NRC) report (1999) entitled *The Assessment of Science Meets the Science of Assessment Standards* identified three main functions of assessment:

1. To monitor educational progress or improvement.
2. To drive changes in practice and policy through accountability.
3. To provide teachers and students with feedback..

It notes that the first two functions pervade schooling while the third remains underdeveloped. Since the release of this report, classroom assessment practices have increasingly integrated formative assessments into curricula.

Multilevel, multitype assessment materials for e-Mission: Operation Montserrat integrate formative assessment as well. However, the added value of these materials derives from the coordination of each of the three functions of assessment detailed in the above NRC report. These materials balance formative and summative functions across different types of assessment practices and at different levels of instructional sensitivity.

Target Learning Standards

Operation Montserrat targets eight learning standards drawn exclusively from the National Science Education Standards (National Research Council, 1995). These eight standards (see Table 1) include content standards for fifth through eighth grades that directly relate to Operation Montserrat and omit all other standards[14].

Table 1.
Target Learning Standards for e-Mission: Operation Montserrat

Content Standard:

- A. Science as Inquiry (pp. 143-148)
    - 1. Abilities necessary to do scientific inquiry
    - 2. Understandings about scientific inquiry
- C. Life Science (pp. 155-158)
    - 1. Structure and function in living systems
    - 4. Populations and ecosystems
- D. Earth and Space Science (pp. 158-161)
    - 1. Structure of the Earth system
- F. Science in Personal and Social Perspectives (pp. 166-170)
    - 1. Personal health
    - 3. Natural hazards
    - 4. Risks and benefits

All assessment materials developed for e-Mission: Operation Montserrat align to these eight

[14] The OM curriculum strongly resonates with some of the explications that follow each target standard, and not with others. This resonance relates well to the functions of a curriculum-oriented exam, and necessarily biases a standards-oriented test.

standards and each activity in e-Mission: Operation Montserrat has been aligned to them as well. Each target standard will be referred to by its unique letter-number combination (e.g., F1 for Personal Health).

The National Science Education Standards also include discussions and enumerated concepts and principles for each standard. The target learning standards in Table 1 omit any and all such explications. This is in accordance with the following caveat tendered together with the standards:

> Following each standard is a discussion of how students can learn that material, but these discussions are illustrative, not proscriptive. Similarly, the discussion of each standard concludes with a guide to the fundamental ideas that underlie that standard, but these ideas are designed to be illustrative of the standard, not part of the standard itself (pp. 6-7).

Table 2.
Alignment Framework

Target Learning Standards	Reading/Activity	Quiz Items	Quiz Answer Explanations	Exam Items	Test Items
A1	• First Step: Form a Team • Your Task (Volcanoes) • Volcano Monitoring Instructions • Volcano Practice Data • Volcano Graphs • Your Task (Hurricanes) • Hurricane Instructions • Evacuation Instructions • Mission Prep: Hurricane Team • Mission Prep: Volcano Team • Mission Prep: Evacuation Team	Q2.2 Q3:2 Q4:2	Q3.2	E:9 E:10	T:13 T:14 T:15
A2	• Earth System Science • How Hurricanes Work • Hurricane Georges • Hurricane Instructions	Q3:2	Q3:2	E:3 E:4	T:4 T:5 T:6
C1	• Yellowstone • Yellowstone Fire: One Year After • Mt. Pinatubo • Mt. Pinatubo: One Year After • Mt. Pinatubo: Five Years After • Your Task (Hurricanes) • Hurricane Georges Hits Puerto Rico • Hurricane Georges: Three Months Later	Q1:2		E:13 E:14	T:19 T:20 T:21
C4	• How They Work (Forest Fires) • Fire Management • Yellowstone • Yellowstone Fire: One Year After • Yellowstone Fire: Six Years After • Hurricane Georges Hits Puerto Rico • Hurricane Georges: Three Months Later	Q1:2 Q1:4 Q3:3	Q1:2 Q1:4 Q4:2	E:15 E:16	T:22 T:23 T:24
D1	• Earth System Science • Forest Fires • Yellowstone Fire: One Year After • Volcanoes • Volcanic Dangers • How Volcanoes Work • Mt. Pinatubo: Five Years After • Volcano Monitoring Instructions • Your Task (Hurricanes) • Hurricanes • How Hurricanes Work • History of Montserrat • Evacuation Instructions • Mission Prep: Maps • Mission Prep: Evacuation Team	Q1:1 Q1:2 Q1:3 Q2:1 Q2:2 Q2:3 Q3:1 Q3:2 Q4:2	Q1:1 Q1:4 Q2:1 Q2:2 Q2:3 Q3:1 Q3:2 Q4:1	E: 1 E:2	T:1 T:2 T:3

Alignment Framework (cont.)

Target Learning Standards	Reading/Activity	Quiz Items	Quiz Answer Explanations	Exam Items	Test Items
F1	• Fire Management • Yellowstone Fire • Mt. Pinatubo • Hurricanes • Hurricane Dangers • Your Task (Montserrat) • Situation Report • Evacuation Instructions • Mission Prep: Volcano Team • Mission Prep: Evacuation Team	Q1:4 Q4:1	Q3:3	E:7 E:8	T:10 T:11 T:12
F3	• Forest Fires • How They Work (Forest Fires) • Volcanic Dangers • How Volcanoes Work • Mt. Pinatubo • Volcano Monitoring Instructions • Hurricanes • How Hurricanes Work • Hurricane Dangers • Hurricane Georges • Evacuation Instructions • Mission Prep: Maps • Mission Prep: Evacuation Team	Q1:1 Q1:2 Q1:4 Q3:2 Q4:1	Q1:1 Q3:2	E:11 E:12	T:16 T:17 T:18
F4	• Team Building • Fire Management • Yellowstone Fire • Yellowstone Fire: One Year After • Volcanic Dangers • Mt. Pinatubo • Mt. Pinatubo: One Year After • Hurricanes • How Hurricanes Work • Hurricane Dangers • Hurricane Georges Hits Puerto Rico • Hurricane Georges: Three Months Later • Your Task (Montserrat) • Newspaper Article • Situation Report • Evacuation Instructions • Mission Prep: Volcano Team • Mission Prep: Evacuation Team	Q1:1 Q1:4 Q2.3 Q3:3 Q4:1 Q4:2	Q1:4 Q2.2 Q3:2 Q3:3 Q4:1 Q4:2	E:5 E:6	T:7 T:8 T:9

Table 3.
Curriculum Outline

Unit	Lesson	Activity	Required Materials	Target Standards
Pretest	X	Administration of the Standards-oriented Test	Standards-oriented Tests (Controlled Document)	A1, A2, C1, C4, D1, F1, F3, F4
Unit 1	1	Introduction to the e-Mission	Introduction Who We Are The Next e-Mission Montserrat Join Us	- - - - -
	2	Applying for the Mission First Step: Form a Team	Application Process Form a Team Earth System Science	- A1 D1, A2
	3	Second Step: Letters of Commitment	Team Building Team Rules Letter of Commitment	F4 - -
	4	Third Step: Writing Résumés	Résumé	-
	5	Fourth Step: Analysis of Yellowstone	Yellowstone Forest Fires Forest Fires How They Work Fire Management	- F3, D1 F3, C1, C4 F4, C4, F1
	6		Yellowstone Yellowstone Fire One Year After Six Years After	C1/C4 F4, F1 C1/C4, F4, D1 C4
	A	Administration of Unit 1 Quiz Unit 1 Feedback Conversation	*Unit 1 Activity-Oriented Quiz* *Unit 1 Answer Explanations*	

Curriculum Outline (cont.)

Unit	Lesson	Activity	Required Materials	Target Standards
Unit 2	7	Volcanoes	Your Task Volcanoes Volcanic Dangers How Volcanoes Work Volcanoes and Montserrat	A1 D1, F3 F3/F4, D1 D1, F3 -
	8	Mt. Pinatubo	Mt. Pinatubo Mt. Pinatubo Eruption One Year After Five Years After	F3, C1, F1, F4 ? F4/C4, C1 D1, C1
	9	Volcano Tracking	Volcano Monitoring Instructions Volcano Practice Data Volcano Graphs	A1, F3, D1 A1 A1
	A	Administration of Unit 2 Quiz Unit 2 Feedback Conversation	Unit 2 Activity-Oriented Quiz Unit 2 Answer Explanations	
Unit 3	10	Hurricanes	Your Task Hurricanes, How Hurricanes Work, Hurricane Dangers	C1, D1, A1 D1, F4, F1, F3 D1, F4/F3, A2 F4/F3, F1
	11	Hurricane Georges	Hurricane Georges Hurricane Georges Hits Puerto Rico Three Months Later	F3, A2 F4, C4, C1 C4, F4, C1
	12	Hurricane Tracking	Hurricane Instructions Hurricane Practice Data Hurricane Tracking Map	A1, A2 - -
	A	Administration of Unit 3 Quiz Unit 3 Feedback Conversation	Unit 3 Activity-Oriented Quiz Unit 3 Answer Explanations	

Curriculum Outline (cont.)

Unit	Lesson	Activity	Required Materials	Target Standards
Unit 4	13	Montserrat	Your Task Montserrat Newspaper Article History of Montserrat Montserrat Fast Facts	F4, F1 - F4 D1 -
	14	Risk Analysis of Montserrat/ Begin Pre-mission Prep	Situation Report Evacuation Instructions Maps	F1, F4 A1, F3, D1, F4, F1 D1, F3
	15	Prepare for e-Mission	Overview of Teams Mission Prep Materials for each team • Hurricane Team • Volcano Team • Evacuation Team • Communication Team	- A1 A1, F4, F1 A1, F3, D1, F4, F1 -
		e-Mission		A1, A2, C1, C4, D1, F1, F3, F4
	A	Administration of Unit 4 Quiz Unit 4 Feedback Conversation	Unit 4 Activity-Oriented Quiz Unit 4 Answer Explanations	
Exam	A	Administration of Final Exam	*Curriculum-Oriented Exams*	A1, A2, C1, C4, D1, F1, F3, F4
	A	Final Exam Feedback Conversations	Completed Curriculum-Oriented Exams Exam Answer Explanations	A1, A2, C1, C4, D1, F1, F3, F4
Posttest	X	Administration of the Standards-oriented Test	Standards-oriented Tests (Controlled Document)	A1, A2, C1, C4, D1, F1, F3, F4

Appendix E. DiSC Instrumentation

Instrument and Inspiration Component or Dimension[1]	Source
Pre-/Postsurvey 1	
Parent's education, race, native language	Adapted from Sloan[2]
Plans for the future: Values, expectations, feelings	Adapted from Sloan[2]: "Your Plans for the Future"
Self-esteem and locus of control	Adapted from Sloan[2]: "Your Opinions"
Perceived social presence	Adapted from Sloan[2]: "Your Opinions"
Friends' values	Adapted from Sloan[2]: "Your Opinions"
Educational plans	Adapted from Sloan[2]: "Your Plans for the Future"
Academic motivation	Adapted from Sloan[2]: "About My Future—Self and Future Expectations"
Mental model: Nature of science	Adapted from Sloan[2]: "About My Future—Job Knowledge"
Mental model: Value of science	Adapted from Sloan[2] "About My Future—Job Knowledge"
Mental model: Science knowledge	Adapted from Sloan[2]: "About My Future—Job Knowledge"
Pre-/Postsurvey 2	
Self-efficacy: Academic	Adapted from Bandura[3]
Self-efficacy: Social	Adapted from Bandura[3]
Self-efficacy: Argumentation	Written by COTF (Reese and Kim) based upon Bandura's self-efficacy guidelines[3]
Mental model: Operation Montserrat-related science career knowledge	Adapted from Sloan[2], written by COTF (Frank): "About My Future—Job Knowledge"
Mental model: Nature of science	Adapted from Lederman and Kuhn[4], written by COTF (Palak, revised by Reese and Kim)
Mental model: Nature of argumentation	Reese and Kim
Source of job knowledge for primary career aspiration	Adapted from Sloan[2]: "About My Future—Job Knowledge"
Aspiration for NASA career	Adapted from Sloan[2]: "About My Future—Job Knowledge"
Pre/Post: Curriculum-oriented Exam *Mental model*	Adapted from COTF Challenger Learning Center assessment instrument[4]
Posttest: Standards-based Test *Mental model*	Items selected from publicly available national and state assessment items—assembled by the University of Georgia multilevel assessment team[5]
Experience Sampling Method Instrument *Flow* and Quality of Experience Variables	Adapted from Sloan study[2]

Notes: [1]Inspiration dimensions are set in italics. [2](Schneider, 1993). [3](Bandura, 2004). [4](Kuhn, 1993; Lederman et al., 2002). [5](Hickey et al., 2004). [6] List of items and sources available upon request. From *Inspiration Brief 2: The DiSC and RoboKids Tools and Labs (Design and Testing),* by D.D. Reese and L. McFarland, 2006, p. 61. Copyright 2006 by Center for Educational Technologies. Used with permission of the authors.

Appendix F. Difficulty and Discrimination Indexes for Each Item within Curriculum-oriented Exam and Standards-based Test

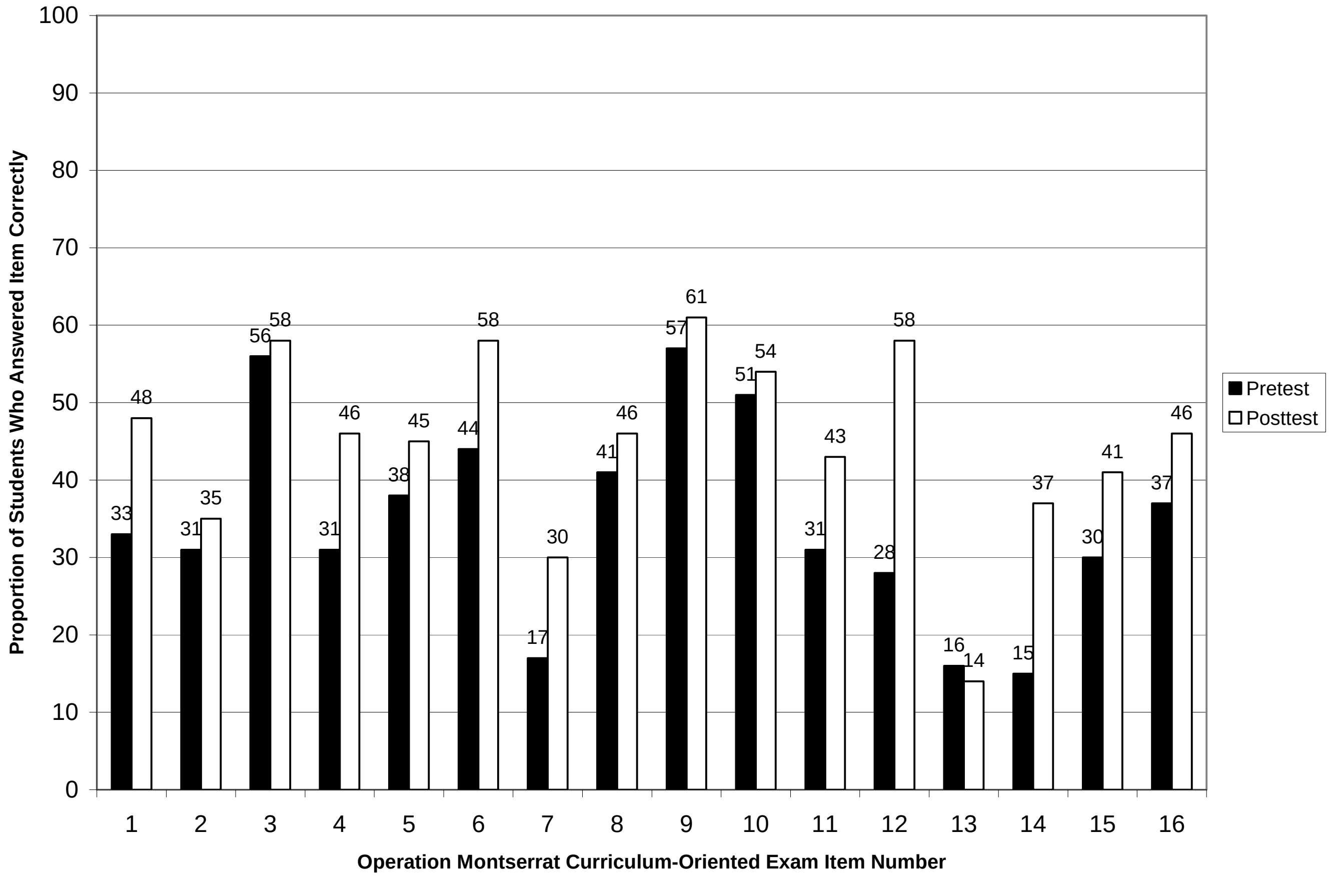


Figure 1. Difficulty Index for Each Curriculum-oriented Exam Item, Pre and Post. The index was calculated across all study participants (n_{pre}=854, n_{post}=757).

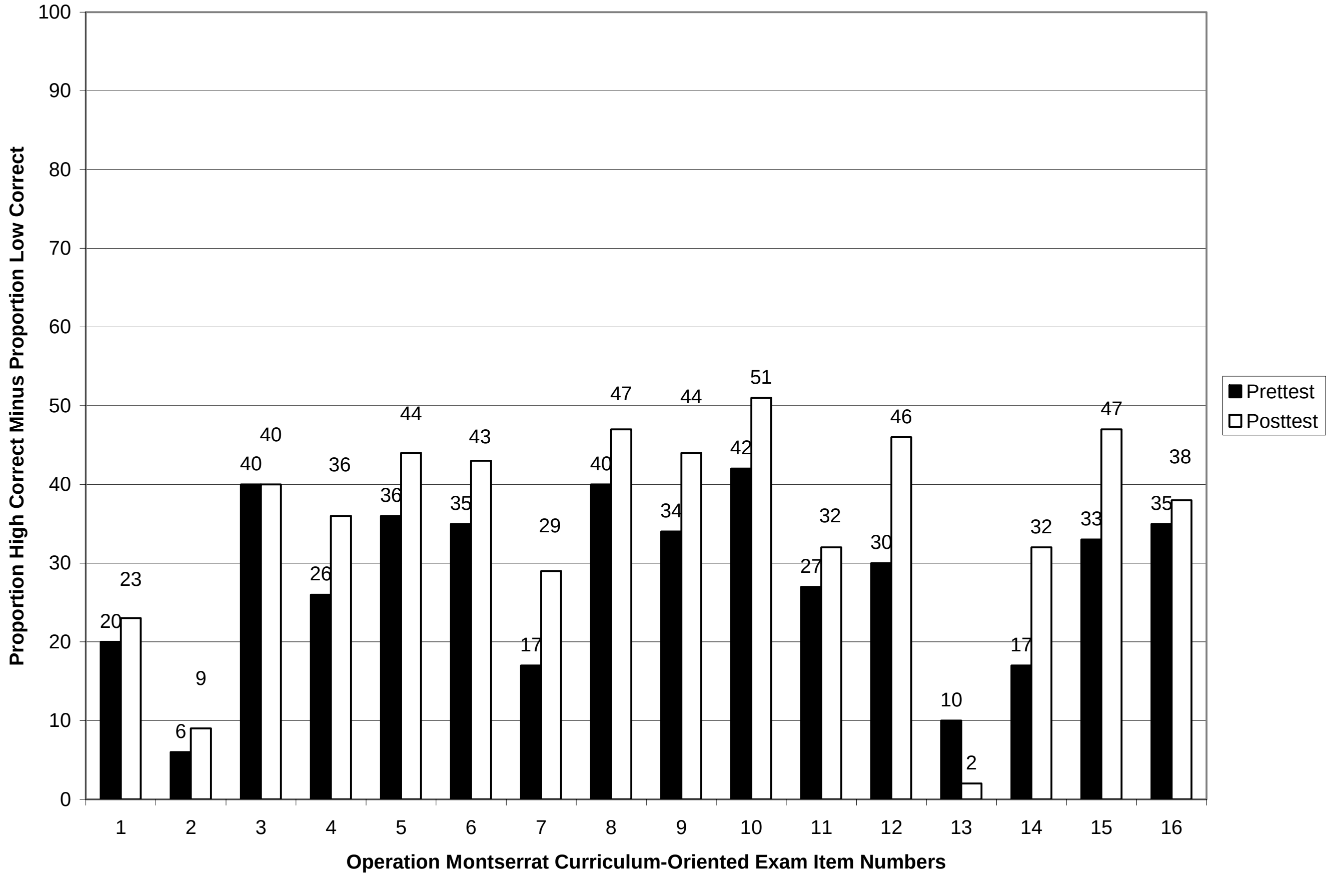


Figure 2. Discrimination Indexes for Each Curriculum-oriented Exam Item, Pre and Post. The index was calculated across all study participants (n_{pre}=854, n_{post}=757).

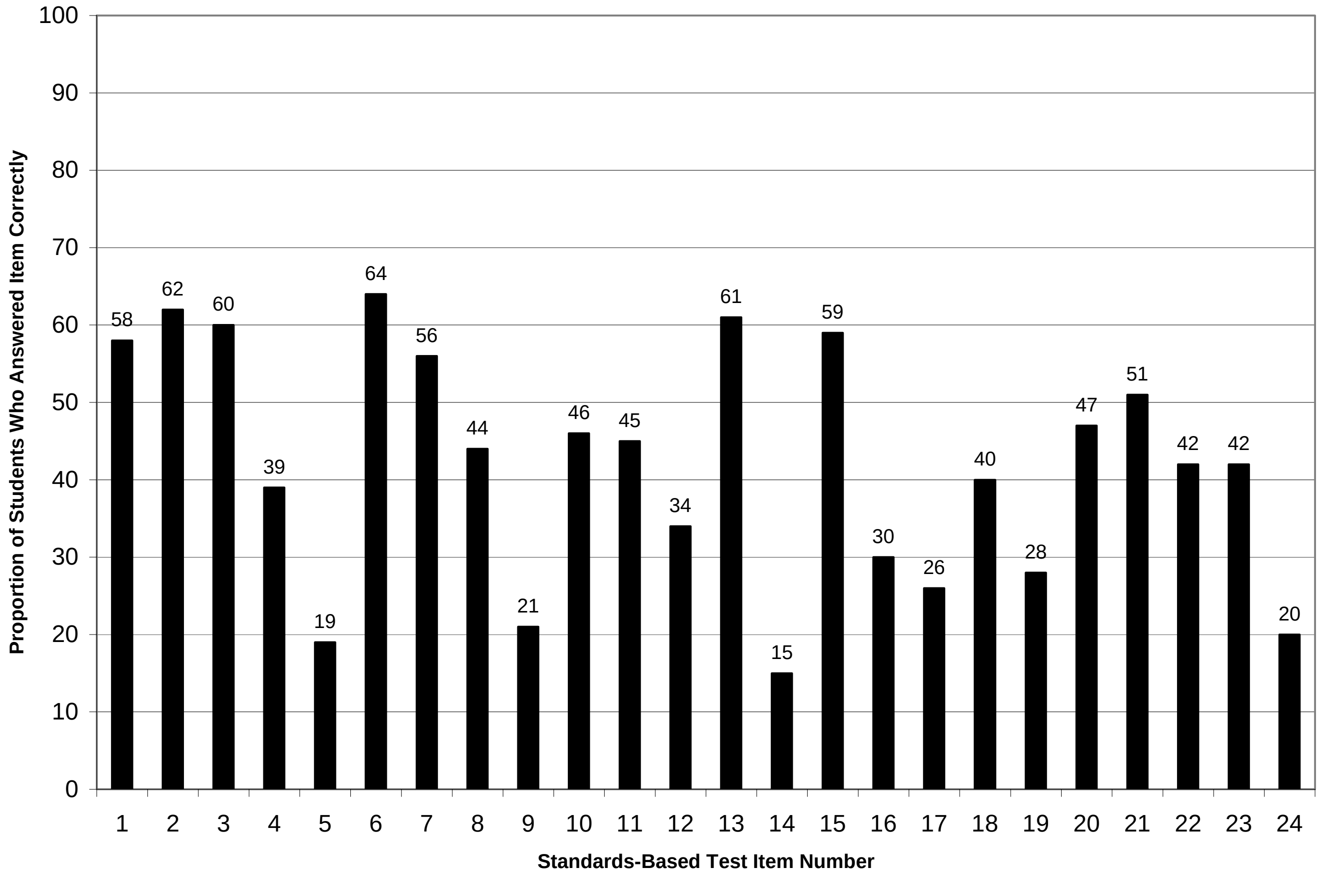


Figure 3. Difficulty Index for Each Standards-based Test (SBT) Pre and Post. The index was calculated across all participants who submitted valid test data (*n*=671); one of these participants was deleted from final participant dataset.

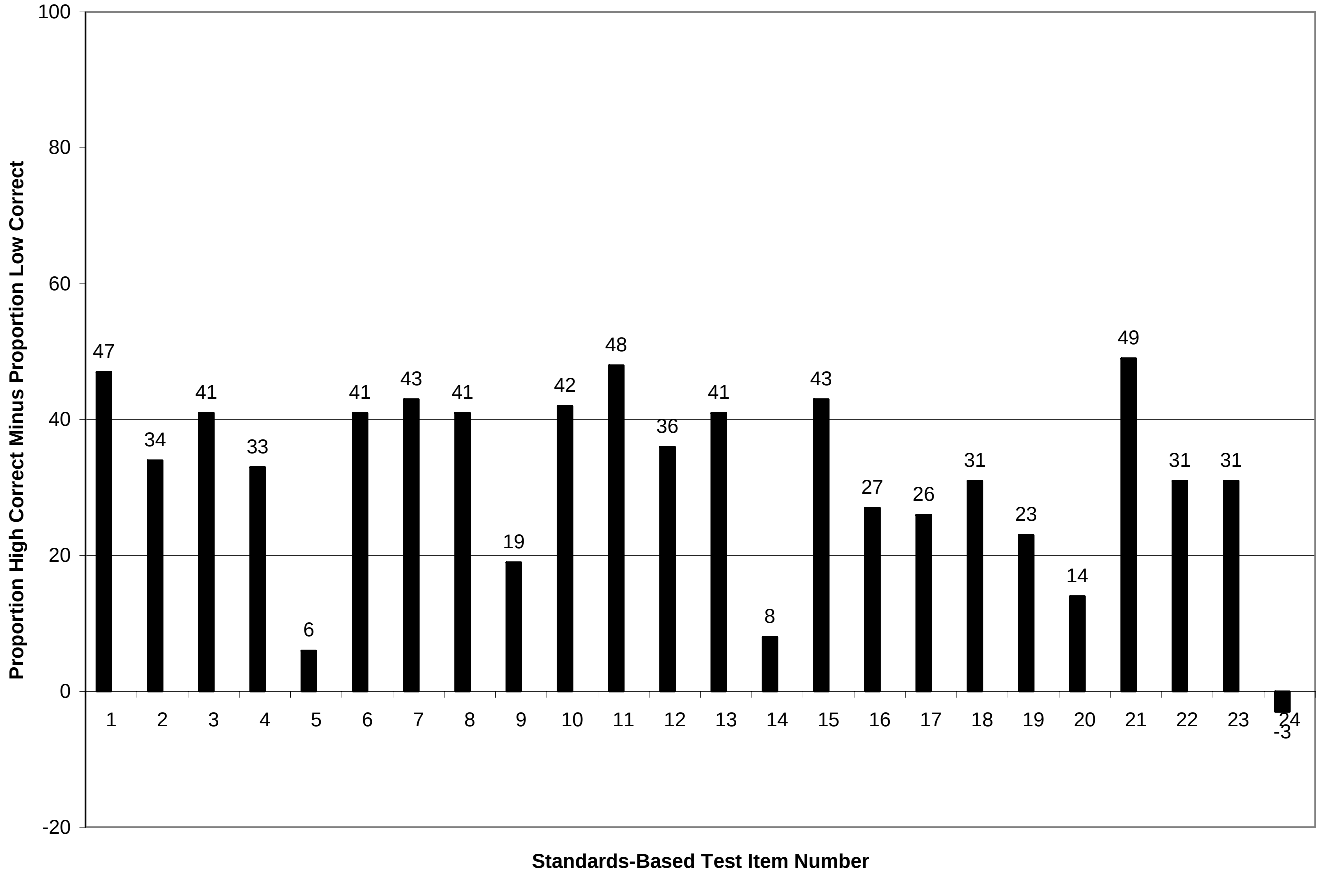
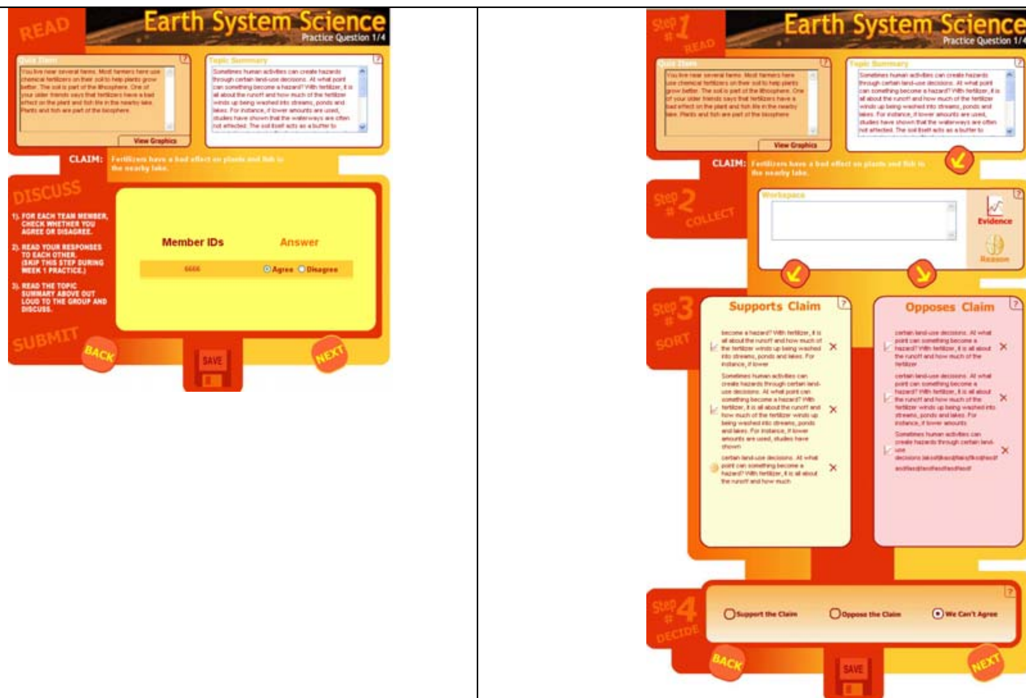


Figure 4. Discrimination Index for Each Standards-based Test Item. The index was calculated across all participants who submitted valid test data (*n*=671); one of these participants was deleted from final participant dataset.

Appendix G. DiSC 2006 Experimental Design

DiSC 2005 analysis can be diagramed from an overall perspective of the study, from the perspective of the tests, or from the perspectives of the ESM data. Figure 1 displays the pretest-postest control group longitudinal design employed over all study instruments. Classes that had been randomly assigned to conditions completed presurveys and a pretest. There were two conditions (see Figure 1): treatment (team use of a DiSC tool that scaffolded team participation in argumentation) and control (team use of a placebo tool with the same look and feel that did not scaffold team argumentation). Students were scheduled to interact with the tool four times during the study (one training session and three tool-usage sessions).



(a) Control: Placebo tool interface

(b) Treatment: DiSC tool interface

Figure 1. The DiSC and placebo tool interfaces.

ESM measures were scheduled daily throughout the study, beginning on study day one. Baseline ESMs provided an indication of states and dimensions of experience before the intervention. The first survey was also administered on day one, followed a second survey on day two and the curriculum-oriented exam on day three. The final activity of the baseline week was tool training. The treatment group was introduced to argumentation via a video and a practice session with the DiSC tool. The control group completed a practice session with the placebo tool.

R	O_A	O_1	X_{1A}	O_2	X_{2A}	O_3	X_{3A}	O_4	X_{4A}	O_5	O_A	O_B
R	O_A	O_1	X_{1B}	O_2	X_{2B}	O_3	X_{3B}	O_4	X_{4B}	O_5	O_A	O_B

Figure 1. Overall Pretest-Posttest Control Group Longitudinal Design.

Notes: O_A = survey and curriculum-oriented exam, O_1 = ESM baseline wave, X_{1A} = DiSC tool training, X_{1B} = placebo tool training, O_2 = ESM wave 2, X_{2A} = DiSC tool session 1, X_{2B} = placebo tool session 1, O_3 =ESM wave 3, X_{3A} = DiSC tool session 2, X_{3B} = placebo tool session 2, O_4 = ESM wave 4, X_{4A} = DiSC tool session 3, X_{4B} = placebo tool session 3, O_5 = e-Mission ESM, O_A = survey and curriculum-oriented exam, O_B = standards-based test.

Data analysis revealed that significant change for the ESM measures of flow (self-reported levels of skill and challenge) occurred only during the e-Mission period. For this reason some of the analyses reported within this brief are limited to differences between baseline and e-Mission data. Limiting the analysis to these two periods reduces the study design to a pretest-posttest control group design (see Figure 2). Collapsing all observations into one signifier, the design looks more like the familiar pretest-posttest control group design (see Figure 3).

R	O_A	O_1	X_A	O_5	O_A	O_B
R	O_A	O_1	X_B	O_5	O_A	O_B

Figure 2. Pretest-Postest Control Group Design.

Notes. O_A = survey and curriculum-oriented exam, O_1 = ESM baseline wave, X_A = DiSC tool training, X_B = placebo tool training, O_5 = e-Mission ESM, O_A = survey and curriculum-oriented exam, O_B =standards-based test.

R	O	X_A	O
R	O	X_B	O

Figure 3. Pretest-Posttest Control Group Design with Observation Types Collapsed into One Symbol.

Notes. O= all presurvey, test, and ESM measures **or** all postsurvey, test, and ESM measures.