Serious Educational Game Assessment

Practical Methods and Models for Educational Games, Simulations and Virtual Worlds

Leonard Annetta and Stephen Bronack (Eds.)

SensePublishers

Serious Educational Game Assessment

Serious Educational Game Assessment

Practical Methods and Models for Educational Games, Simulations and Virtual Worlds

Edited by

Leonard Annetta
George Mason University, Fairfax, VA, USA

Stephen C. Bronack
Clemson University, Clemson, South Carolina, USA

SENSE PUBLISHERS
ROTTERDAM/BOSTON/TAIPEI

A C.I.P. record for this book is available from the Library of Congress.

ISBN: 978-94-6091-327-3 (paperback)
ISBN: 978-94-6091-328-0 (hardback)
ISBN: 978-94-6091-329-7 (e-book)

Published by: Sense Publishers,
P.O. Box 21858,
3001 AW Rotterdam,
The Netherlands
http://www.sensepublishers.com

Printed on acid-free paper

TABLE OF CONTENTS

EDITORIAL

In a reality of accountability and subsequent assessments along with the prolific use of the term 21st century skills, there seems to be a decoupling of sorts. On one hand, we are educating our K-12 teachers and students to infuse 21st century technology into their teaching and learning repertoire; yet, on the other, we continue holding them accountable for performance on 19th century tests and other antiquated measures. Serious Educational Games, Simulations, and Virtual Worlds are common platforms by which today's students collaborate, communicate, become entertained and in many cases learn; albeit often times not in a traditional school setting. It has been our contention for some time that these environments hold the key to providing authentic learning activities and performance assessments. The educational holy grail of sorts, Serious Educational Games, Simulations and Virtual Worlds finally are realizing their true potential in support of teaching and learning.

School connectivity is increasing and cloud computing is becoming a realty in K-20 education. Database technology is affordable and the world's cyberinfrastructure is growing by the day. Data mash-ups also are becoming mainstream and the mashing of these varying technologies are finding their way into formal and informal learning environments through one delivery mechanism: 3-dimensional virtual learning environments such as Serious Educational Games, simulations and virtual worlds.

We hope that this volume will not be viewed as the panacea of assessment in 3-dimensional spaces; rather, it is a starting point for exploring another dimension of effective use of these innovations. From a book Len edited in 2008 also published by *Sense* entitled, *Serious Educational Games,* it became apparent that there was a need to share with the broader community how these ever-evolving technologies can be used to assess student learning, efficacy and the like. The one chapter missing from the first book was a chapter on assessment. Although much of the chapters were written about work done on a National Science Foundation funded project, we neglected to share how we empowered the technology as a research and assessment mechanism.

The timeline unfolded quite rapidly from there when a colleague, David Gibson, asked Len to present with him at the Society for Technology in Teacher Education (SITE) conference in 2009. David was conducting some interesting assessments in one of his projects as well. What became apparent from that one-hour session was that there was a great interest in this line of work. Further, the *Games and Simulation* Special Interest Group at the same conference expressed both interest and concerns about assessment in games, simulations and virtual worlds.

When we began talking about the work we've been doing and how assessment seems to be the missing link, this edited volume seemed like a natural progression. We solicited chapters from a wide variety of researchers, independent contractors and game developers to contribute to this book. What follows are a series of chapters that share, in large part, what is being done in these spaces as it pertains to assessment.

We do not offer this collection as an all-encompassing work. Instead, we suggest that the reader will find in these pages a compendium of some of the most creative, relevant, and useful accounts of how pioneers in Serious Educational Games, Simulations, and Virtual Worlds are bridging the gap between providing innovative spaces for learning and assessing their effectiveness. Each of the chapters in this book details a process, or a product, or a theoretical construct that can serve as a model, a tool, or a foundation upon which to base your own effort and exploration into 21st century assessment. The array of work included in this book is broad, and grounded in disciplines as varied as art education, entertainment provision, and automobile construction. Yet, the tie that binds it together is an explicit commitment to quality assessment that exploits the unique characteristics of these emerging media within the bounds of time-tested constructs such as rigor, validity, and reliability.

Feedback from the community is always welcome, as we recognize that we have barely scratched the surface. As technology continues to evolve and research methodologies become more prevalent in Serious Educational Games, Simulations and Virtual Worlds, we hope those working in 3D spaces leave their silos and share all they know with the real world. We hope you enjoy this collection and thank you for your interest in this body of work.

We’d also like to acknowledge the Marta Klesath, Rebecca Cheng and Beth Bagwell for their help in reviewing the wonderful contributions to this book.

FORWARD

Len Annetta, lead author of this book, was the best man in my virtual wedding on April 22, 2010. In fact, you can watch Len, my lovely bride, and our wedding here on the archive: http://www.ustream.tv/channel/setsermarcus2010-virtual-wedding. Moreover, you could have participated in the virtual cake cutting, toasted a virtual cocktail, and/or danced with guests at the reception in second life: http://slurl.com/secondlife/North%20Carolina%20Virtual/197/124/25

If those reading this book, do not think that gaming, simulations, and e-learning have penetrated our face-to-face society on a massive scale, then you need to put down your stones and fire and take heed to not get eaten by the "Changeasaurus" among us that has become virtual worlds.

As Chief Executive Officer of the North Carolina Virtual Public School (NCVPS), www.ncvps.org, I have been "failing forward" for the past three years in the e-learning space. Yet, with over 16 years of public school experience, I fully understand the challenges of virtual content, context, and touch points for assessing student learning, attitudes, and outcomes in face-to-face institutions as well as in virtual environments.

The challenges and risks are immense, but the rewards that are being obtained by the chapter authors in this work are profound. As more is published in this arena, I continue to be fascinated by the potential for learning through simulations and serious games. Not only do we at NCVPS expect our employees to play and immerse themselves in innovative solutions, we have a virtual structure whereby employees self-regulate they're working hours, time with their families, and time they spend on creating and innovating over the net. Such autonomy in our workforce captures the very essence of many of the pioneers in this book. These authors are part of the next generation of thinkers who eat "barriers for breakfast" when it comes to open, immersive, and emergent approaches for learners in serious games and virtual worlds.

As the authors explore case studies and topics such as constraints and limitations in serious games, they take the reader on a game path of their own. Reviewers of this work will achieve understandings, navigate impossible impasses, and enlist the help of multiple chapter authors to make sense of how to meet objectives from studying this work.

Not unlike a serious game or the simulated virtual classrooms I lead at NCVPS, Cafés in Second Life (SL) can be an informal meeting place for teaching about digital visual culture and for hosting learning events for the public. Yet what seems so obvious to those of us in the virtual environment is not always the case with the larger k-20 community. How does one begin to comprehend the pedagogy around an environment that theoretically works like trying to build a lesson plan around multiple people who attended the Luvre on the same day? It is to this point, that game play is vastly different than the teaching and learning paradigms we have become accustomed to in our pre-kindergarten through higher education systems.

Learners truly can choose as the Florida Virtual Public School motto suggests anytime, anywhere, any path, any pace by which to complete their learning journey

(FLVS, 2010). In fact, Florida's recent game creation of Conspiracy Code: http://www.flvs.net/areas/flvscourses/ConspiracyCode/Pages/default.aspx, could be even more cutting edge by the pioneers in this book via their assertions that the learning is from adjustable, user centered created, and able to be immersive and self-directed in its' discovery for the learner through serious games.

Therefore, my interest in being involved in this work begins with assessing student learning. The wonderful part about serious games and virtual worlds is that you are not arrested for killing the patient in the allied health seminar. Or, as you fail Arabic diplomacy in the simulated environment for language acquisition, a covert operative does not sell you out as a just pegged Americano. The point here is that learners are given the opportunity as Toffler says, "to learn, unlearn, and relearn" through serious gaming and simulations. Moreover, they are allowed to collaborate with others and tap the wisdom of crowds to solve common problems within the simulation. For example the military has known for some time that pilots need repeated simulations to perform under duress. Therefore, they reach out to companies like http://www.lockheedmartin.com/sts/virtual/index.html to design simulations and worlds for everything from training to strategic visioning.

So why is this work slow to catch on in the education space? Well, institutions of higher education would say they are already fueling the innovation in this space as seen by the number of universities who have a presence in www.secondlife.com or www.activeworlds.com. Yet, it has been in my domain of k-12 where serious games and virtual worlds have had trouble getting off the ground. The reasons for this are manifold in terms of technology infrastructure, human capital, and cost, but the largest issues are the dependency on paper pencil or online assessments as the primary means of assessing how students learn.

As NCVPS is a trailblazer in this space for k-12 education, we have begun to set up student learning "field trips" in our Russian Courses using a company out of North Carolina -http://www.americanri.com/tools/poweru.php where students practice running their own cell phone store through a series of serious game modules written and spoken in Russian. This work is significant in that it assesses student competencies in language, math for the cell phone business plan, and cultural skills in the environment. Yet, the best part of the effort is that students are not assessed in the traditional "averaging" sense of student grades. Rather, serious games and virtual worlds allows for "mastery learning" in that learners do not progress until they have mastered and/or collaborated for mastery with an array of resources, docents, instructors, and/or exemplars within the virtual world.

The irony of the malaise in k-12's foray into serious gaming and simulations is that the leaders, policy makers, and educators see this environment every day in the news and laud its impact. For example, Captain Sully Sullenburg did not land that plane on the Hudson River last year, because he practiced "real" water landings over and over. Quite the contrary, the Captain was highly trained in the simulation environment until his skills became honed to execute such maneuvers in the real world. This approach has to become more of the norm in k-20 environments so that we feed these sets of 21st Century Skills up to higher education institutions for the purpose of simulated learning games and virtual world outcomes for all learners.

The authors in this work provide many points of research around attitudes in virtual worlds both from the learner and instructor perspectives. Attitudes in the k-20 environment range from fear, to lack of competence, to job security issues. Therefore, it is important to help educators learn in a safe space like www.usdlc.com, www.inacol.org, and/or at www.iste.com, where educators can truly play, practice, and make determinations around serious games and simulations and the research and results behind them. For example, at NCVPS we began our journey into virtual worlds by organizing around professional development in second life through a conversation between my avatar and Michael Horne of *Disrupting Class* Fame: http://www.bookosphere.net/michaelhorn/.

Allowing educators opportunities to log on, take care of Maslow type needs like clothing, and ultimately learn professional development techniques begins to bridge the gap as to how they might transfer that knowledge to student learning in the classroom. However, where the authors in this book excel is when they illustrate discussions on student outcomes. Throughout the work, you'll read about student participation levels, academic excellence, and time spent on learning in serious games and virtual worlds.

Yet, we are indeed at a true crossroads in education as far as the level of penetration in serious games and virtual worlds in the education and the economic sector. The cost of constructing these worlds is not minimal. As vendors continue to approach me on a daily basis at NCVPS, all promise to build me the perfect Algebra I learning game for just under a cool million. This model is not sustainable and makes open source worlds at the elementary level like www.webkinz.com and www.clubpequin.com far more scalable in terms of the skills we teach emergent learners to navigate, learn, and produce outcomes in the virtual world. As long as we are discussing scale, the case of http://www.farmville.com/ comes to mind.

Once a routine "hide" for me on face book from the besiege of all my friends has now become equivalent to one of the largest countries in the world. Why has the phenomenon of Farmville achieved such scale? It's free to play. Learners of all ages access the content. It's fun to play, and it allows you to socially connect to others. This social connection is something that is often missed by those who downplay the effectiveness of serious games and virtual worlds. As avatars become far more life like and convincing, www.facestation.com, and http://www.xbox.com/en-US/live/projectnatal/, the trick with tapping virtual worlds and serious games is how will we integrate multi-media, game play, and free open source software in such a way that the value add of this learning paradigm is no longer titled serious games, virtual worlds, or e-learning – yet is simply titled learning.

We are already seeing http://www.uopeople.org/ as the first tuition free university. So, how long before we'll see the Farmville equivalent here? Will the authors in this book contribute their expertise, develop mobile applications for the work, support the student learning, and/or serve as consumers? The answer for tomorrow's learning is that the lessons in this book contribute to the knowledge base of how we move from where we are to where we need to be – ubiquitous access via immersive learning across all devices and platforms. To achieve such an ambitious goal, Virtual worlds must get lighter and cheaper without compromising quality. In an effort to take the collective expertise in this work to scale, the lessons learned from these pioneers fall out into three categories of service that will help share the future of serious games and virtual worlds.

First, the classic environment of virtual worlds is already in place. Large bandwidth hogs and hard to run engines on school networks must give way to lighter engines like www.teleplace.com to begin the serious games and virtual world journey with few barriers around implementation.

Second, the modular nature of serious games and virtual worlds must be condensed to small clees and applications that people can load to the computer, mobile device, and/or training platform to have short, powerful experiences in serious games and virtual worlds for course refreshers, compliance requirements, and innovative solutions like playing www.urgentevoke.com on your phone to socially collaborate around the solutions of tomorrow. The World Bank sponsored this solution, it is free, and allows multiple people to collaborate in a virtual world "light" environment.

Finally, we are in transition in this country in the k-20 space around "blended learning." In the context of what the authors contend in this work, professional learning, play, development and student exercises all take place along side of face to face lectures, sit and get instruction, and bad exemplars in e-learning. Therefore, the thought leadership in this book provides key building blocks for tomorrow's future in serious games and virtual worlds. For my part, I am honored to be asked to contribute to the body of this work, and my wedding in second life will not be the last time I flirt with blending my world with the virtual one.

Bryan Setser
CEO, North Carolina Virtual Public School
www.ncvps.org

DIANE JASS KETELHUT

1. ASSESSING GAMING, COMPUTER AND SCIENTIFIC INQUIRY SELF-EFFICACY IN A VIRTUAL ENVIRONMENT

INTRODUCTION

“She’s here!” the whispers spread through the room. I had arrived to give my volunteer science lesson to my daughter’s fifth-grade class. These children loved science. For them, this subject was the high point of the day. Yet, as a former veteran secondary science teacher, I know that this is not true of older children. Somewhere between elementary and high school, many students disengage from science, partially because they develop a belief that they can’t succeed in science (Clark, 1999; Farenga & Joyce, 1998).

Bandura (1977) coined the term “academic self-efficacy” for this belief that one can succeed in learning a particular type of content. This phenomenon, recognized by teachers, is quantified in the *National Assessment of Educational Progress* (National Center for Educational Statistics, 2000). An item-stem measuring self-efficacy in this large survey asserts “all can do well in science if they try.” 82% of fourth-graders agreed with this statement; however, only 64% of eighth-graders and a mere 44% of twelfth-graders agreed (NCES, 2000).

In this chapter, I discuss the development, piloting and revision of a new instrument for measuring pre/post the academic self-efficacy of students. The initial motivation for the creation of this new instrument came from an unsuccessful search for an instrument to measure self-efficacy in an NSF-funded project designed to investigate the motivational effects of a multi-user virtual environment (MUVE) on the science achievement of middle-school students. However, individual sections of the new instrument can contribute in general to the more-finely discriminated measurement of self-efficacy, particularly for middle grades students. The following section provides a background and context for the development of the new instrument. Next, the research design used in the data collection is described, outlining the key features of the sample, measures, procedures and data-analyses. Finally, findings and conclusions are presented.

BACKGROUND OF RESEARCH

Bandura (1977) defined “self-efficacy” as the belief that one can successfully perform certain behaviours. He hypothesized that individuals obtain information about their self-efficacy in four ways. First, students’ own performances affect their self-efficacy: Students who successfully graph data will feel more confident when next

L. Annetta and S. C. Bronack, (eds.), Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations and Virtual Worlds, 1–18.

asked to graph data. Second, students' vicarious experiences affect their self-efficacy: When a student sees a peer successfully graphing data, she may feel more confident when asked to graph data herself. Third, students' self-efficacy can be affected by others' verbal persuasion: a teacher may persuade a student that she can successfully graph data, and thus she approaches the next graphing task confidently. The fourth factor is emotional arousal: for example, a student's confidence in approaching a graphing exercise inversely depends on his level of anxiety induced by that assignment. All of these experiences can affect self-efficacy either positively or negatively. Past failures lower a person's perception of his or her abilities to perform certain tasks successfully, while past successes raise their perceptions (Pajares, 1997).

Research on the impact of academic self-efficacy on student behaviour and learning in the classroom is dense. A detailed review of this literature is beyond the scope of this paper (see for example, Pajares, 1996; Schunk, Pajares, Wigfield & Eccles 2002). In short, levels of self-efficacy impact perseverance, engagement and success in the classroom; and affect how students view complexity and performance (Collins, 1984; Lent, Brown, & Larkin, 1984; Pajares, 2000; Pajares, 1995; Bandura, 1986; Zimmerman & Bandura 1994). Further, self-efficacy in a particular domain predicts career interest in that field (Lopez & Lent, 1992).

This last finding is of particular interest for science educators. A number of reports recently have pointed to concern over the number of students interested in a scientific career (National Science Foundation, 2001; Grigg, Lauko, & Brockway, 2006; Committee on Prospering in the Global Economy, 2007). Other research indicates that this interest begins early. For example, Tai and colleagues (2006) found a strong correlation between student interest in a scientific career while in eighth grade and ultimately having one.

One method for improving interest in a scientific career is to expose students to scientific inquiry. The National Science Education Standards define scientific inquiry as encompassing the ways that scientists investigate and explain the natural world (NRC, 1996). Gibson and Chase (2002) found that middle-school students participating in an inquiry-based summer program demonstrated increased interest in a scientific career. Given the research showing that higher levels of self-efficacy predict for scientific career interest, it is tempting to hypothesize that the relationship between scientific inquiry experiences and career interest is mediated by changes in self-efficacy. Unfortunately, research into this relationship is difficult to conduct well in typical K-12 classrooms primarily because of the paucity of good classroom-based scientific inquiry (Nelson & Ketelhut, 2007).

As a result of the issues related to integrating scientific inquiry into the classroom, some researchers interested in exploring classroom-based inquiry are using computers to support this (Nelson & Ketelhut, 2007; deJong, 2006). In particular, in the last decade, virtual environments have been investigated as a medium for scientific inquiry-based curricula (for a review, see Nelson & Ketelhut, 2007). However, little research has been conducted on whether these have the same impact on career interest as do other forms of scientific inquiry, nor has there been much research on the impact of virtual environments on self-efficacy (Ketelhut, 2007; Nelson & Ketelhut, 2007).

Dede & Ketelhut (2003) used a graphical multi-user virtual environment (MUVE) to engage students in a collaborative science-inquiry learning experience and measured self-efficacy in general science pre-post. Students conducted scientific investigations in a virtual historical town—populated by themselves, Smithsonian artifacts and computer agents—in order to decipher a pattern of illness sweeping through the virtual community. Findings suggest that the MUVE treatment was motivating and enhanced student science learning, particularly for lower-ability students. In addition, the self-efficacy (at least in general science) of students in the MUVE treatment improved over the course of the study compared to a control group that received a parallel curriculum not involving participation as a scientist in a virtual world (Nelson, Ketelhut, Clarke, Bowman & Dede, 2005).

Since self-efficacy was defined in the late 1970s, many researchers have designed instruments to measure the construct. Some of these instruments are framed generally, while others are designed for administration in specific settings and intellectual contexts. Midgley and colleagues (2000) developed a general measure of self-efficacy for administration to K-12 students. Within the instrument, item-stems are written generically: "I am certain I can master the skills taught in class this year." Conversely, Miltiadou & Yu (2000) created an online instrument designed specifically to measure the self-efficacy of college students using online communications. Here, item-stems were specific: "I would feel confident reading messages from one or more members of the synchronous chat system."

In previous research on science learning, a general measure of self-efficacy was used to evaluate students' confidence in their ability to learn science (Dede & Ketelhut, 2003). Unfortunately, Bandura (1986, 1997) and others (Miltiadou & Yu, 2000; Pajares, 1996; Smith and Fouad, 1999) have argued that it is not sufficient to measure self-efficacy globally; this construct must be measured in a context-specific way. Therefore, Bandura suggests that measures of self-efficacy be designed to be as specific to the task being performed as possible. While some researchers have cast doubt on this (Bong, 1996), the strength of the relationship between self-efficacy and performance appears to weaken when more global instruments are used to measure it (Pajares, 1995).

In investigating self-efficacy in multi-user virtual environments, it would be important to identify and separate out the various components that might contribute to changes in student engagement. MUVEs are a form of gaming and require students to manipulate an avatar and communicate with peers using an instant messaging format. While there are numerous self-efficacy measures for some aspects of technology and science, currently there are no context-specific instruments available for measuring self-efficacy in these gaming technological aspects, nor for middle-school students using an inquiry process to learn aspects of experimental scientific investigation. A more specific instrument is needed to measure academic self-efficacy reliably and validly in research on scientific inquiry, within a specific technology experience—a MUVE—and for use in the broader study of MUVEs. Such an instrument with independently designed subscales could also find use in the larger context of game-related or simulation-based science curriculum, as well as technology-based learning experiences in general.

Although some researchers may use unpiloted survey instruments in their studies, conducting a measurement pilot to establish the reliability and validity of the data of a new instrument in the intended empirical context is considered crucial by most (Litwin, 2003; Popham, 1981). For example, in creating their online self-efficacy instrument for college students, Miltiadou & Yu (2000) estimated internal consistency reliability, as measured by Cronbach's alpha, to establish the precision of their instrument. Midgley et al (2000) also estimated the Cronbach's alpha reliability of their instrument.

In addition, self-efficacy researchers have adopted a variety of methods for demonstrating the validity of the data of their instruments. Midgley et al (1998) provided evidence of their instrument's construct validity by examining the estimated correlations between scores on their instrument and scores on instruments measuring other closely-related constructs, as indicated in the literature. To create a measure of general self-efficacy Jinks and Morgan (1999) demonstrated the convergent validity of their instrument by estimating the correlations of self-efficacy scores obtained using their own instrument with students' self-reported grades. Finally, they demonstrated the content validity of their instrument by having experts review and comment on its content.

In this chapter, I report on the development of an instrument—*Self-Efficacy in Technology and Science* (*SETS*)—for measuring the self-efficacy of a middle-school student in technology and science. In the development of this new instrument, I address three research questions. My first question asks: What is the internal consistency reliability of each subscale of *SETS* and what contribution does each item make to the subscale internal consistency reliability? My second question asks: Do the subscales of the *SETS* instrument demonstrate both content and construct validity? Finally, I ask: Do the modified subscales of the *SETS* instrument maintain their internal consistency reliability when implemented with a large urban middle school population?

RESEARCH DESIGN

Site

I implemented the measurement pilot in 2 middle schools, one in suburban California and the other in suburban Massachusetts. I chose these sites based on their interest in the project. In the second phase of the study, I implemented the modified subscale in middle schools from four large urban districts throughout the Midwest and Eastern United States.

Sample

I used a convenience sample of 98 students in the measurement pilot (the sample for the second phase was over 2000 students and will be described later). I attempted to ensure that the sample contained representatives of the different genders, abilities and ages present in the general middle school population. My final sample fulfilled most, but not all, of these criteria. Sampled students were nearly evenly split by gender, 54% male and 46% female. The split by grade was less balanced, with the majority (58%) of the students in seventh grade, a sizeable proportion (33%) in eighth grade, and the remaining 8% split between fifth and sixth grades. The ability of these students was slightly above average, as suggested by their average science grade of B+.

Measures

I designed the initial version of the *SETS* instrument to measure self-efficacy in four sub-contexts: (a) *Computer Use*, (b) *Videogaming*, (c) *Synchronous Chat Use*, and (d) *Inquiry Science* (see Appendix A). These sub-contexts represent the content and process areas affected by technology-based inquiry science projects situated in MUVEs. As indicated above, virtual environments are seen as 'games' by the children using them. While participating in a virtual environment such as ours, students apply videogame skills and interact with teammates via synchronous chat while conducting scientific inquiry. I formulated the sections of the *SETS* instrument on *Videogaming, Synchronous Chat Use* and *Computer Use* to allow assessment of the effects of self-efficacy in these areas on learning while taking part in a virtual environment-based curriculum. Each of the four sections contained 15–16 item-stems and Likert-type response statements. I modeled the content of these items after items on similar surveys in the literature (Miltiadou & Yu, 2000; Enochs, 1993; Koul, 1999; Baldwin et al, 1999; Ruess, 2001).

I worked with a team of five experts who were members of the larger NSF-funded MUVE-based project's consultant panel to evaluate the draft instrument. These consultants were all experts in gaming, curriculum design and scientific inquiry. Each expert evaluated the instrument for content, clarity and appropriateness. I then modified the item-stems based on their feedback prior to implementation.

Procedures of the Measurement Pilot

The order of the statements in the four subsections were randomized and administered to sample children online using *Survey Monkey* (http://www.surveymonkey.com/), with supervision provided by the participating teachers. Before the test was administered, teachers read a prepared statement to all students, describing the purposes of the project and how to proceed. Students then responded to sixty-one self-efficacy statements before answering ten questions that sought demographic data. Students worked from individual computers, inputting their answers to the questions as they appeared on the screen. They were required to answer all questions in a section before moving to the next set. Students finished the survey in a single class period.

Data Analysis

To address the first research questions on reliability for each *SETS* subscale, I conducted both an item analysis to estimate the internal consistency reliability of the SETS scale, including the reliability of its subscales, and a principal components analysis to estimate the contribution that each item made to each subscale. First, I estimated alpha reliability prior to any modification of the subscale. I followed this with a principal components analysis of the student responses to items on the subscale to determine which items should be deleted or if the subscale should be split into other subscales based on how the individual items clustered. Finally, I re-estimated the alpha reliability of the modified scale.

For the second research question on validity, I collected evidence for the content validity of the instrument and its subscales by asking experts to examine and critique items to determine if the instrument measured the intended content. I then supported the construct validity of each subscale by estimating the correlations between children's responses on the instrument with three other variables, measuring: (a) their hours of use of technology per day, (b) their gender and (c) their overall grade in their most recent science course; all theorized to be related to self-efficacy. Remember that, as noted earlier, students can improve their self-efficacy in several ways; in essence, success breeds success (Bandura, 1977). As a result, I hypothesized that students who use the various technologies more would have a higher self-efficacy in those areas, since their increased practice would result in more opportunities for success than those who play less or not at all. Therefore, if the *SETS* subscale scores are valid, I would expect a positive correlation between each *SETS* subscale score and its associated technology.

Further, research on gender and videogame use indicates that males play videogames more frequently than females and feel more competent in using them than females do (Turkle, 1995; Murray, 1997). Thus, if males have more practice at videogaming, then their self-efficacy in videogaming should be higher. Therefore, if the *SETS* subscale scores on V*ideogaming* are valid, I would expect a positive correlation between scores on that subscale and being male.

Finally, as mentioned previously, Bandura (1977) posits that past successes raise a person's perceptions of their own ability. Therefore, those with greater past successes might be expected to have a higher sense of their own competence. Assuming that student grades on the most recent science course are a measure of past success in the classroom, I would expect that students with higher science grades, and thus greater success in class, would have higher self-efficacy in inquiry science. I would therefore expect a positive correlation between grades and self-efficacy in inquiry science if the scores for this *SETS* subscale were valid.

Lastly, I re-estimated the alpha reliability of the newly modified scales with implementations in four large urban school districts to confirm the reliability in a larger sample that also included a broader audience base.

FINDINGS

RQ1. What is the Internal Consistency Reliability of the Subscales of the SETS Instrument?

In Table 1, I summarize the reliability analysis for the *SETS* subscales. In column 1, I list the five subscales as they were formatted initially, and, in columns 2 and 3, I list

Table 1. Number of items and estimated internal consistency reliability for each subsection of SETS (n=98)

	At start			*At end*	
Initial self-efficacy subsection	*Number of item- stems*	*Reliability*	*Final self-efficacy subsection*	*Number of item- stems*	*Reliability*
Computer Use	16	.86	General Computer Use	11	.80
			Problem-solving Computer Use	.5	.79
Videogaming	15	.94	Videogaming	.8	.93
			Computer Gaming	.5	.84
Synchronous Chat Use	15	.92	Synchronous Chat Use	10	.92
Inquiry Science	15	.90	Inquiry Science	12	.90

the number of items and estimates of internal consistency reliability for each subscale. My initial reliability analysis suggested that I should split two of the initial subscales into two further subscales and that I refine two of the remaining subscales. I split the C*omputer Use* subscale into the G*eneral Computer Use* and the *Problem-solving Computer Use* subscales. I split the *Videogaming* subscale into the *Videogaming* subscale and the *Computer Gaming* subscale. I modified the subscales *Synchronous Chat Use* and *Inquiry Science*. In column 4 of Table 1, I show the final subscales and in columns 5 and 6, I show the number of items and the reliability estimates for these final subscales.

As can be seen in the first row, column 2 and 3 of Table 1, the *SETS* subscale on *Computer Use* self-efficacy initially consisted of 16 items with an initial internal consistency reliability estimate of .86. I conducted a principal components analysis of these 16 items. The principal components analysis suggests that items on this subscale formed two distinct and interesting clusters. “Learning how to use a computer is not hard for me” is an example of an item-stem from the first cluster; “It is hard for me to use a computer to do my homework” is an example from the second cluster. This analysis thus indicates that these items are measuring different constructs; therefore, I split the initial subscale into two. The first new subscale, *General Computer Use,* contained eleven of the original items that formed cluster 1, and the second, *Problem-solving Computer Use,* contained the remaining five items, found in cluster 2. I re-estimated the internal consistency reliability for each of the two new subscales (see columns 5 and 6, rows 1 and 2 of Table 1). The estimated reliability for the eleven items representing *General Computer Use* was only slightly decreased to .80, while the five remaining items representing *Problem-solving Computer Use* was lower at .79, possibly reflecting the decrease in items.

As stated earlier, I also split the *Videogaming* subscale into two subscales as a result of the reliability analysis. The *SETS* subscale on *Videogaming* self-efficacy initially consisted of 15 items with an initial internal consistency reliability estimate of .94 (see third row, column 2 and 3 of Table 1). The principal components analysis suggests that items on this subscale also formed two distinct and interesting clusters, one representing gaming on computers and the other videogaming. Thus, I split this subscale into two subscales as well. The first new subscale, *Videogaming,* contained eight of the original items, including "I am certain that I can master any videogame I play"; the second, *Computer Gaming,* contained five of the original items, including "I can figure out most computer games." I re-estimated the internal consistency reliability for each of the two new subscales (see columns 5 and 6, rows 3 and 4 of Table 1). The estimated reliability for the eight items representing *Videogaming* was still high at .93, while the five items representing *Computer Gaming* was lower at .84.

The *Synchronous Chat Use* subscale initially consisted of 15 items with an internal consistency reliability estimate of .92. The principal components analysis suggested that ten of the items on this subscale consist of a single cluster, with the remaining five items being somewhat independent of the rest. An example of an item from the tight cluster is "If I need help, I can talk to my friends online." I modified the *Synchronous Chat Use* subscale to consist of the ten closely clustered items, and re-estimated the internal consistency reliability (see columns 5 and 6, row 5 of Table 1). The estimated reliability for the ten items representing *Synchronous Chat Use* remained the same at .92.

As can be seen in the sixth row, column 2 and 3 of Table 1, the *SETS* subscale on *Inquiry Science* also initially consisted of 15 items with an initial internal consistency reliability estimate of .90. The principal components analysis suggested that items on this subscale form a single cluster of twelve items. "I can use graphs to show what I found out in my experiment" is an example of an item-stem in this cluster. I re-estimated the internal consistency reliability for this modified version of the *SETS* subscale on *Inquiry Science* (see columns 5 and 6, row 6 of Table 1). The estimated reliability for the twelve items representing *Inquiry Science* was unchanged at .90.

RQ2. Do the Subscales of the SETS Instrument Demonstrate Content and Construct validity?

Do the subscales of the SETS instrument show content validity. Once I had confirmed the final make-up of the subscales based on the reliability and component analysis, I showed the survey again to the team of experts for evaluation of content validity. They did agree that each section appeared to demonstrate content validity. Several experts affirmed the separation of the *Computer Gaming* and *Videogaming.* However, one concern that the team of experts indicated was the possible overlap between the *General Computer Use* section and the section on *Problem-solving Computer Use.* In the first, there is a statement: "I can learn how to write papers on any computer." Since the reliability analysis places this into the cluster with other items in the *General Computer Use* self-efficacy subscale, I assumed that this was measuring whether students could use a word processing program. However, there are two statements in the second cluster of items in the *Problem Solving Computer Use* subscale that sound very similar: “It is hard for me use a computer to do my homework” and “Even if I try very hard, I cannot use the computer as easily as paper and pencil.” One expert questioned the validity of saying these statements are measuring different concepts from “I can learn how to write papers on any computer.” However, other experts on the team did not see this as an issue. More investigation into this subscale is needed to confirm these results.

Do the subscales of the SETS instrument demonstrate construct validity. To establish the construct validity of an instrument usually involves examining estimated correlations among scores on the subscales and other variables theorized to be related to them. No measure can be judged construct valid in perpetuity; evidence for construct validity is continually accumulated over time (Cronbach & Meehl, 1955). For this first attempt, I examined the estimated correlations of responses to each self-efficacy subscale with selected variables measuring hours of technology use, gender, and science grade. Table 2 shows the correlations between the six self-efficacy subscales in row 1 with the five related variables in column 1. The highest correlation(s) for each subscale is indicated in bold.

I hypothesized that if the *SETS* subscale on *Videogaming* was valid, then the more hours that students played videogames, the higher would be their self-efficacy. In row 2 of Table 2, I display the estimated correlations between hours of videogaming per day and scores on the six new subsections of the SETS scale. Notice that

Table 2. Estimated correlations of theoretically-related variables and scores on the six new subsections of SETS (n=98)

	Video-gaming	*Computer gaming*	*Scientific inquiry*	*Synchronous chat*	*Computer use*	*Computer problem-solving*
Hours of videogaming	**.58*****	.37***	.12	.11	.04	.04
Hours of computer gaming	**.47*****	**.47*****	.15	.02	.20*	.20*
Hours of synchronous chat use	.08	.04	.12	**.63*****	.06	.09
Being male	**.59*****	.46***	.31**	.03	.03	.08
Science grades	.02	.16	**.44*****	.08	.23*	.18~

~ p<.10 * p<.05 ** p<.01 *** p<.001

scores on the SETS subscale on Videogaming are highly correlated with hours of videogaming, thereby supporting the validity of the Videogaming subscale. Notice also that the estimated correlation between hours of videogaming and self-efficacy subscale scores for Videogaming is larger than the estimated correlation between hours of videogaming and self-efficacy subscale scores for Computer Gaming. I separated these two subscales from the original single subscale as a result of my reliability analysis, and the difference in the two correlations reported here suggests that each subscale measures different constructs, reinforcing the reliability analysis.

To demonstrate evidence of validity of the *Computer Gaming* and *Synchronous Chat Use* subscales, I hypothesized that the more hours students played computer games or chatted online, the higher their self-efficacy subscore would be for *Computer Gaming* and *Synchronous Chat Use,* respectively. The estimated correlations between hours of computer gaming and synchronous chat and children's scores on the six new self-efficacy subscales in rows 3 and 4, respectively. As hypothesized, hours of computer gaming are highly correlated with the student scores on the SETS subscale on Computer Gaming. In this situation, the estimated correlation between hours of computer gaming and student scores on the Computer Gaming subscale are identical to the estimated correlation between hours of computer gaming and scores on the Videogaming subscale. However, remember that I found that the estimated correlation between hours of videogaming and scores on the Computer Gaming subscale were not the same as the estimated correlation between hours of videogaming and the scores on the Videogaming subscale. One possible interpretation of this differential result is that while videogamers are as likely as computer gamers to play computer games, the converse is not as likely (Jones, 2003). Further, synchronous chat use is highly and only correlated with students' scores on self-efficacy in synchronous chat as measured by the SETS subscale, Synchronous Chat Use. This provides strong evidence for the validity of this subscale.

In addition to hours of use, I hypothesized that being male would be correlated with scores on the Videogaming and Computer Gaming self-efficacy subscales. Row five of Table 2 shows the estimated correlations between gender and scores on the six new self-efficacy subscales. Since gender is a dichotomous predictor, correlations estimated with gender act as t-tests on the averages of subscale scores for boys and girls. Since scores on the Videogaming and Computer Gaming subscales are highly correlated with being male, this indicates that the average self-efficacy scores of boys are statistically significantly higher than for girls on these two subscales, which supports construct validity for these two sections. In addition, construct validity is also demonstrated because the average scores for boys and girls do not differ for the Synchronous Chat Use, Computer Use and Problem-solving Computer Use subscales where the literature does not indicate any differences. However I did not hypothesize that, on average, males would have a moderately higher self-efficacy score on the Inquiry Science subscale than girls. Nonetheless, that there is such a correlation is supported by findings in the literature that report that middle school is the time that girls begin to question their abilities as science learners (Clark, 1999; Farenga & Joyce, 1998). Therefore, this finding, of concern to science educators, also demonstrates construct validity for the Inquiry Science subscale.

I obtained a final demonstration of the validity of the *SETS* subscales by estimating correlations between the students' grade in science and their scores on the six new subscales, shown in row six of Table 2. The high magnitude of the estimated correlation between science grade and self-efficacy in inquiry science supports validity for the subscale, Inquiry Science, as hypothesized. A correlation between science grades and gaming or synchronous chat was not expected nor was it found, further supporting construct validity. The small correlation between the self-efficacy subscore for Computer Use and Problem-solving Computer Use and science grade was not hypothesized, but seems logical, as computer use has become a mainstream tool for helping students with research, projects and homework. Those students confident enough to use the computer would be likely to produce better work and achieve higher grades.

RQ3. Do the Modified Subscales of the SETS Instrument Maintain their Internal Consistency Reliability when Implemented with a Large Urban Middle School Population?

To provide stronger reliability information for the newly modified SETS instrument, several of the subscales were incorporated into the pretests for students who participated in the afore-mentioned MUVE project. Over 300 urban seventh and eighth grade students responded to the items in the new self-efficacy in Videogaming and self-efficacy in Computer Gaming subscales. These middle-school students come from two major metropolitan areas, one in New England and the other in the Midwest and they represent a different and more diverse population from the original pilot study. The internal consistency reliability estimates for these two subscales continues to be high with only slight modifications from the original small study: .86 for the Videogaming subscale and .89 for the Computer Gaming subscale.

The self-efficacy in Inquiry Science subscale has been used extensively in the virtual environment project with over 2000 students. These primarily middle-school students come from urban and suburban schools across the Midwest, Southeast and Northeast. With this larger population, the internal consistency reliability estimate for this subscale has varied little from the original: .86. These further studies support the initial decisions made regarding the estimates of internal consistency reliability.

CONCLUSION

I created the initial *SETS* instrument with four subscales on science and technology self-efficacy for ultimate use in a research project on multi-user virtual environments. After completing a reliability analysis, I was persuaded that I needed to split two of the subscales further and condense the other two. The final *SETS* instrument now consists of six subscales, each with between five and twelve items, and with the reliability estimates ranging from .79 to .93 (see Appendix B). In the reported validity analysis, I examined relationships between use of technology, gender, and science grades and scores on each subscale and affirmed that each subscale was measuring what was intended. In the process of doing this, I discovered that middle-school girls had lower self-efficacy scores in using gaming media and in conducting scientific inquiry. This is an important finding that must be considered in future design work.

The use of emerging technologies, such as MUVEs, applies principles well-understood by the entertainment industry to engage students in their own learning. Once students perceive themselves as competent, their effort and perseverance will increase. As a result of this, students will be able to be challenged with more complex material which may make them more competitive globally. However, to evaluate whether these technologies are having the desired effect, we need appropriate assessment measures. This development of a measure of self-efficacy is a first step in that process.

REFERENCES

Baldwin, J. A., Ebert-May, D., & Burns, D. J. (1999). The development of a college biology self-efficacy instrument for nonmajors. *Science Education*, *83*(4), 397–408.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman & Company.

Bong, M. (1996). *Perceived similarity among tasks and generalizability of academic self-efficacy*. Paper presented at the annual meeting of the American Educational Research Association, New York, April 8–12, 1996.

Clark, J. V. (1999). *Minorities in science and math*. Columbus: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Collins, J. L. (1984). *Self-efficacy and ability in achievement behavior*. Unpublished Dissertation/Thesis, Stanford University.

Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology, National Academy of Sciences, National Academy of Engineering, Institute of Medicine. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: The National Academies Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

de Jong, T. (2006). Technological advances in inquiry learning. *Science*, *312*, 532–533.

Dede, C., & Ketelhut, D. (2003). *Designing for motivation and usability in a museum-based multi-user virtual environment*. Paper presented at the American Educational Research Association, Chicago.

Enochs, L. G. (1993). The development and partial validation of microcomputer utilization in teaching efficacy beliefs instrument in a science setting. *SchoolScience & Mathematics*, *93*(5), 257–263.

Farenga, S. J., & Joyce, B. A. (1998). Science-related attitudes and science course selection: A study of high-ability boys and girls. *Roeper Review*, *20*(4), 9.

Gibson, H., & Chase, C. (2002). Longitudinal impact of an inquiry-based science program on middle school studentsâ€™ attitudes toward science. *Science Education*, *86*, 693–705.

Grigg, W., Lauko, M., & Brockway, D. (2006). *The Nation's report card: Science 2005* (NCES 2006–466). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Jinks, J., & Morgan, V. (1999). Children's perceived academic self-efficacy: An inventory scale. *The Clearing House, 72*(4), 224–230.

Jones, S. (2003). *Let the games begin: Gaming technology and entertainment among college students* (Report). Washington, DC: Pew Internet & American Life Project.

Ketelhut, D. J. (2007). The Impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in river city, a multi-user virtual environment. *Journal of Science Education and Technology*, *16*(1), 99–111.

Koul, R. (1999). An analysis of the reliability and validity of personal internet teaching efficacy beliefs scale. *Electronic Journal of Science Education*, *4*(1).

Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology, 31*, 356–362.

Litwin, M. S. (2003). *How to assess and interpret survey psychometrics* (2nd ed., Vol. 8). Thousand Oaks, CA: Sage Publications.

Lopez, F. G., & Lent, R. W. (1992). Sources of mathematics self-efficacy in high school students. *The Career Development Quarterly*, *41*, 3–12.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., & Anderman, et al. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: U. of Michigan.

Midgley, C., Kaplan, A., Middleton, M., Maehr, M., Urdan, T., Anderman, L., et al. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology, 23*, 113–131.

Miltiadou, M., & Yu, C. H. (2000). *Validation of the Online Technologies Self-Efficacy Scale (OTSES)* (Publication no. ED445672). Retrieved January 15, 2004.

Murray, J. H. (1997). *Hamlet on the Holodeck*. Cambridge, MA: MIT Press.

National Center for Educational Statistics. (2000). *The nation's report card*. Retrieved October 12, 2003, from http://nces.ed.gov/nationsreportcard/

National Research Council. (1996). *National science education standards: Observe, interact, change, learn*. Washington, DC: National Academy Press.

National Science Foundation. (2001). *Characteristics of Doctoral Scientists and Engineers in the United States, 2001* [Internet]. National Science Foundation. Retrieved March 19, 2004, from http://www.nsf.gov/sbe/srs/nsf03310/pdf/tab6.pdf [].

Nelson, B., & Ketelhut, D. J. (2007). Designing for real-world inquiry in virtual environments. *Educational Pyschology Review*, *19*(3), 265–283.

Nelson, B., Ketelhut, D. J., Clarke, J., Bowman, C., & Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology*, *45*(1), 21–27.

Pajares, F. (1995). *Self-efficacy in academic settings.* Paper presented at the American Educational Research Association, San Francisco.

Pajares, F. (1996). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence*. Retrieved February 21, 2004, from http://www.emory.edu/EDUCATION/mfp/aera2.html

Pajares, F. (1997). Current directions in self-effiacacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 1–49). Greenwich, CT: JAI Press.

Pajares, F. (2000). *Schooling in America: Myths, mixed messages, and good intentions* (Report). Cannon Chapel, GA: Emory University.

Popham, W. J. (1981). *Modern educational measurement*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Ruess, K. (2001). [Email sent 2001, November].

Schunk, D. H., Pajares, F., Wigfield, A., & Eccles, J. (2002). The development of academic self-efficacy. In Anonymous (Ed.), *Development of achievement motivation*. San Diego, CA: Academic Press.

Smith, P. L., & Fouad, N. A. (1999). Subject-matter specificity of self-efficacy, outcome expectancies, interests, and goals: Implications for the social-cognitive model. *Journal of Counseling Psychology*, *46*(4), 461–471.

Tai, R., Liu, C., Maltese, A., & Fan, X. (2006). CAREER CHOICE: Enhanced: Planning early for careers in science. *Science*, *312*(5777), 1143–1144.
Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. New York: Simon & Schuster.
Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*, 845–862.

Diane Jass Ketelhut
Temple University

APPENDICES

APPENDIX A

Initial Subsections of Sets

Videogaming-15 items.

I am certain that I can master any videogame I play.
No matter how hard I try, videogames are too difficult for me to learn. (reverse item)
Even when I try hard, I don’t do well at playing videogames. (reverse item)
I can only succeed at the easiest videogame. (reverse item)
No matter how hard I try, I do not do well when playing computer games. (reverse item)
Videogames are hard to figure out for me even when I try. (reverse item)
Even when I try hard, learning how to play a new videogame is complicated for me. (reverse item)
When playing in a simulation game, I only do well for short periods of time. (reverse item)
I can keep winning at computer games for a long time.
I can learn how to play any computer game if I don’t give up.
I am sure that I can learn any videogame.
I am very good at building things in simulation games.
I can do well at even the most challenging videogame.
I can figure out most computer games.
I am sure that I can succeed at playing videogames.

Synchronous chat-15 items (modified from Miltiadou & Yu, 2000).

If I need help, I can talk to my friends online.
I can create and use a new username on chat programs whenever I want.
I can use chat programs online.
When I need an answer to a question, I cannot find it by ‘chatting’ with a friend online. (reverse item)
I can show how I am feeling online by using happy or sad faces.
I can chat online with people.
Chat rooms are too difficult for me to figure out. (reverse item)
I can talk with my friends using shortcuts like LOL or BRB.
I am very good at carrying on conversations with my friends online.
It is hard for me to talk with my friends online. (reverse item)
I find it difficult to carry on many conversations with friends online at the same time. (reverse item)
I am sure that I can join a chat room and talk with several people at the same time.
I can use Instant Messenger.
No matter how hard I try, I cannot make chat programs work. (reverse item)
Learning how to use instant messenger is hard for me. (reverse item)

Generic computer usage-16 items (modified from Miltiadou & Yu, 2000; Enochs et al, 1993; Koul, 1999).

I know the steps necessary to use the computer to create a presentation.
I can open and close software programs on a computer.
I can turn a computer on and off.
It is hard for me to use a computer to do my homework. (reverse item)
It is hard for me to look for answers to questions on the Internet. (reverse item)
Learning how to use a computer is not hard for me.
No matter how hard I try, I cannot learn how to use computers (reverse item).
I find it difficult to learn how to use the computer. (reverse item)
I know how to sign on to the Internet.
I know I can find a specific website if I have the address (URL).
Whenever I can, I avoid using computers (reverse item).
When using the Internet, I usually have trouble finding the answers I am looking for. (reverse item)
Even if I try very hard, I cannot use the computer as easily as paper and pencil (reverse item).
I can find information on the web by using a search engine.
I can learn how to write papers on any computer.
If I need to learn how to do something on a computer, I can do it.

Inquiry science processes-15 items (modified from Baldwin et al, 1999; Ruess, 2001).

No matter how hard I try, I cannot figure out the main idea in science class. (reverse item)
When I do not understand something in science, I know where to find help.
I can write an introduction to a lab report.
I can use graphs to show what I found out in my experiment.
When I do a science fair project, it is hard for me to come up with a question to research. (reverse item)
It is hard for me to write a report about an experiment. (reverse item)
I know how to use the scientific method to solve problems.
It is hard for me to look at the results of an experiment and tell what they mean. (reverse item)
When I do an experiment, it is hard for me to figure out how the data I collected answers the question. (reverse item)
When I do my work in science class, I am able to find the important ideas.
Once I have a question, it is hard for me to design an experiment to test it. (reverse item)
I can design an experiment to test my ideas.
I have trouble figuring out the main ideas of what my science teacher is teaching us. (reverse item)
I can tell the difference between observations and conclusions in a story.
It is easy for me to make a graph of my data.

Typical Format of Survey:

		STRONGLY DISAGREE	**DISAGREE**	**SOMEWHAT AGREE**	**AGREE**	**STRONGLY AGREE**
66	It is easy for me to make a graph of my data.	○	○	○	○	○
67	Learning how to use instant messenger is hard for me.	○	○	○	○	○

APPENDIX B

Final version of SETS by subsection

SE on science inquiry—12 items

Internal consistency reliability: .90

I can write an introduction to a lab report.
I can use graphs to show what I found out in my experiment.
It is hard for me to write a report about an experiment.
I know how to use the scientific method to solve problems.
It is hard for me to look at the results of an experiment and tell what they mean.
When I do an experiment, it is hard for me to figure out how the data I collected answers the question.
When I do my work in science class, I am able to find the important ideas.
Once I have a question, it is hard for me to design an experiment to test it.
I can design an experiment to test my ideas.
I have trouble figuring out the main ideas of what my science teacher is teaching us.
I can tell the difference between observations and conclusions in a story.
It is easy for me to make a graph of my data.

Self-efficacy on Videogaming—8 items

Internal consistency reliability: .93

I am certain that I can master any videogame I play.
No matter how hard I try, videogames are too difficult for me to learn.
Even when I try hard, I don’t do well at playing videogames.
Videogames are hard to figure out for me even when I try.
Even when I try hard, learning how to play a new videogame is complicated for me.
I am sure that I can learn any videogame.
I can do well at even the most challenging videogame.
I am sure that I can succeed at playing videogames.

Self-efficacy on computer gaming—5 items

Internal consistency reliability: .84

No matter how hard I try, I do not do well when playing computer games.
I can keep winning at computer games for a long time.

I can learn how to play any computer game if I don't give up.
I am very good at building things in simulation games.
I can figure out most computer games.

Self-efficacy on General Computer use—11 items

Internal consistency reliability: .80

I know the steps necessary to use the computer to create a presentation.
I can turn the computer on and off.
Learning how to use a computer is not hard for me.
I know how to sign on to the Internet.
I know I can find a specific website if I have the address (URL).
I can open and close software programs on a computer.
No matter how hard I try, I cannot learn how to use computers.
I find it difficult to learn how to use the computer.
Whenever I can, I avoid using computers.
I can learn how to write papers on any computer.
If I need to learn how to do something on a computer, I can do it.

Self-efficacy on Problem-Solving Computer Use—5 items

Internal consistency reliability: .79

It is hard for me to look for answers to questions on the Internet.
It is hard for me to use a computer to do my homework.
When using the Internet, I usually have trouble finding the answers I am looking for.
Even if I try very hard, I cannot use the computer as easily as paper and pencil.
I can find information on the web by using a search engine.

Self-efficacy on synchronous chat use—10 items

Internal consistency reliability: .92

If I need help, I can talk to my friends online.
I can create and use a new username on chat programs whenever I want.
I can use chat programs online.
When I need an answer to a question, I cannot find it by 'chatting' with a friend online.
I can show how I am feeling online by using happy or sad faces.
I can chat online with people.
I am very good at carrying on conversations with my friends online.
It is hard for me to talk with my friends online.
I am sure that I can join a chat room and talk with several people at the same time.
I can use Instant Messenger.

JOHN NIETFELD AND LUCY R. SHORES

2. SELF-REGULATION WITHIN GAME-BASED LEARNING ENVIRONMENTS

INTRODUCTION

Digital games have the potential to provide an ideal environment for students to “learn how to learn.” However, this potential remains as of yet untapped (Squire, 2006). It is no secret that today’s youth have become mesmerized by computer games and gaming consoles. The motivating factors inherent in games grab the attention of youth compelling them to play the same game for hours at a time. Moreover, youth do not seem deterred by the time required to conquer the steep learning curve necessary to succeed in complex, open-ended gaming environments. For these reasons, game-based learning environments (GBLEs) have garnered increasing attention in the educational research community. However, serious games, or games that are used for purposes other than entertainment, are still not common in educational settings. By juxtaposing the captivating nature of games with educational content, it seems reasonable to assume students would find similar motivation to conquer GBLEs that could discreetly yield educational gains. Furthermore, these environments provide promising contexts in which to study self-regulated learning (SRL) due to the complex yet autonomous settings that they afford.

While research developments in GBLEs and in SRL are both in their adolescent stages they are growing at a rate unrivaled by most other contemporary research topics. Gee (2003) and others have emphasized the importance of learning a new literacy through games but only recently have educational psychologists acknowledged the potential of digital games as a platform in which to study SRL. SRL has been assigned many definitions, but we choose Schraw’s (2007) conceptualization of SRL as the process of managing one’s learning, which includes planning, goal setting, strategy implementation, summarizing, and monitoring one’s progress. Models of SRL are composed at the broadest levels of strategic, metacognitive, and motivational components (Zimmerman, 2000). Currently, there is a shift in efforts to understand self-regulation not only in traditional learning environments but also in computer-based learning environments (CBLEs) that are capable of providing a dynamic and adaptive context for learning (Graesser, McNamara, & VanLehn, 2005).

Digital games have the potential to promote autonomous learning within a social constructivist framework that emphasizes ill-structured problem solving. However, this poses significant challenges for research designs due to the complexity of GBLEs and the additional processing demands required of the user (Lajoie & Azevedo, 2006; Schraw, 2007). Moreover, existing empirical research in the area is somewhat

L. Annetta and S. C. Bronack, (eds.), Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations and Virtual Worlds, 19–42.

lacking with regard to clear evidence of the impact of GBLEs on targeted learning outcomes and therefore is at risk of being disregarded as "motivational fluff" (O'Neil, Wainess, & Baker, 2005, p. 456).

In order to provide sophisticated models of SRL in GBLEs, carefully designed studies are needed not only to examine the relationships between key variables within the proposed models but also to determine the most effective means by which to integrate SRL into GBLEs. The purpose of this chapter is to give some direction for future research. First, we will provide background in SRL-related research along with a proposed multilevel model of SRL. We will then review findings from three existing bodies of research that may inform future studies examining SRL within GBLEs (see Figure 1). These areas of research include studies of SRL in traditional learning environments, intelligent tutoring systems (ITS), and game-based learning studies. We believe that integrating findings from these three existing bodies of research will not only inform future studies but will also contribute to more sophisticated and fully developed models of SRL in GBLEs. Following this review, we look to the future by presenting areas of needed research in GBLEs organized by the levels of SRL as presented in the proposed model. Thus, our goal for the conclusion of this chapter is to provide some critical questions for future research as well as possible avenues of investigation that have emerged from this literature.

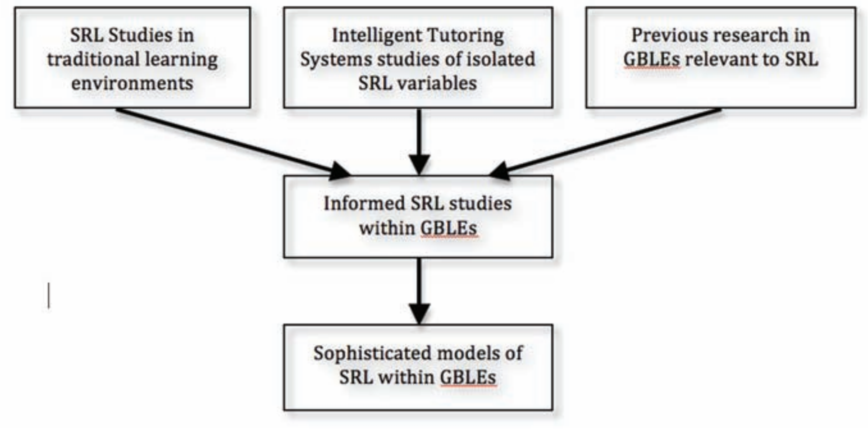
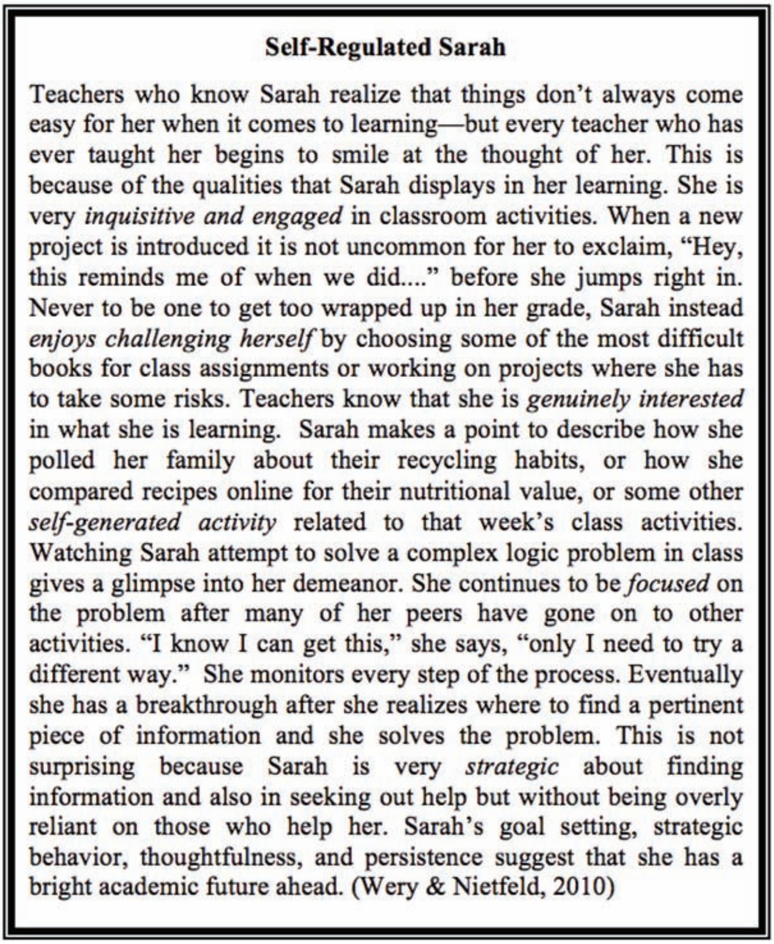


Figure 1. Progression of research of SRL in game-based learning environments.

WHAT DOES IS MEAN TO BE A SELF-REGULATED LEARNER AND WHY IS THIS CRITICAL?

Describing Self-regulated Learning

Self-regulation is an expanding research topic that has attracted attention from numerous bodies of literature (Boekaerts, Pintrich, & Zeidner, 2000). SRL is expressed by beliefs, motives, strategies, and reflective processes that allow learners to autonomously direct their own learning (see Figure 2 for example). SRL requires the knowledge of many cognitive strategies (e.g., summarizing, elaboration, etc.) in addition to strategies related to setting up an environment conducive to learning and the use of appropriate help seeking. Metacognition assists in managing strategy use and knowledge and has been defined as a framework consisting of knowledge of cognition and regulation of cognition (Baker & Brown, 1984; Schraw & Moshman, 1995). Knowledge of cognition refers to individuals' declarative, procedural, and conditional knowledge about their thinking and memory processes. Regulation of cognition involves intentional control of one's cognition, memory, or learning and



Self-Regulated Sarah

Teachers who know Sarah realize that things don't always come easy for her when it comes to learning—but every teacher who has ever taught her begins to smile at the thought of her. This is because of the qualities that Sarah displays in her learning. She is very *inquisitive and engaged* in classroom activities. When a new project is introduced it is not uncommon for her to exclaim, "Hey, this reminds me of when we did...." before she jumps right in. Never to be one to get too wrapped up in her grade, Sarah instead *enjoys challenging herself* by choosing some of the most difficult books for class assignments or working on projects where she has to take some risks. Teachers know that she is *genuinely interested* in what she is learning. Sarah makes a point to describe how she polled her family about their recycling habits, or how she compared recipes online for their nutritional value, or some other *self-generated activity* related to that week's class activities. Watching Sarah attempt to solve a complex logic problem in class gives a glimpse into her demeanor. She continues to be *focused* on the problem after many of her peers have gone on to other activities. "I know I can get this," she says, "only I need to try a different way." She monitors every step of the process. Eventually she has a breakthrough after she realizes where to find a pertinent piece of information and she solves the problem. This is not surprising because Sarah is very *strategic* about finding information and also in seeking out help but without being overly reliant on those who help her. Sarah's goal setting, strategic behavior, thoughtfulness, and persistence suggest that she has a bright academic future ahead. (Wery & Nietfeld, 2010)

Figure 2. Example of a self-regulated student.

manifests itself in regulatory behaviors such as planning, monitoring, and evaluation in the learning and problem solving process. In addition, motivational factors such as intrinsic motivation, interest, goal directedness, goal-orientation, task value, self-efficacy, and attributions, help drive SRL.

Recent research shows that all learners use self-regulatory processes to some degree, but exemplary learners display awareness of strategic relations between regulatory processes and outcomes and modify their strategies adaptively (Ellis & Zimmerman, 2001). According to self-regulated learning theory, learners rely on systematic internal monitoring and feedback systems (Butler & Winne, 1995). In fact, Pintrich (2000) described one common feature across all models of SRL to be the potential for control assumption. In other words there is potential for learners to regulate, monitor, and control their cognition but this does not mean that this will occur at all times. A critical component in this process is the *motivation* for the learner to engage these control processes. Moreover, the quality of the self-regulatory skills that students employ depends in part on underlying beliefs that students hold about themselves and the process of learning more broadly. More specifically, high achieving students possess more metacognitive awareness and engage in more self-regulatory behavior than low achieving students (Hacker, Bol, Horgan, & Rakow, 2000). Thus, SRL can be viewed as a multivariate continuum of strategic, metacognitive, and motivational variables.

A Model of Self-regulated Learning

We situate our investigation within a theoretical model that recognizes the contextual nature of SRL and combines an orientation from information processing and social cognitive theories (see Figure 3). Within the model there are three levels of SRL: the task level, the domain level, and the dispositional level. The task level represents a simplified synthesis of three predominant SRL theories (Winne & Hadwin, 1998; Pintrich, 2000, Zimmerman, 2000) encapsulated within a learning context. It includes three phases that learners progress through on any given task: anticipatory, enactment, and self-reflection. The *anticipatory* phase includes the process of defining the task, assessing available resources and constraints, and determining task goals. In addition, learners approach the task with particular beliefs about knowledge, goal orientations, interest, and judgments of task value and self-efficacy that will impact their engagement. Finally, learners bring to the task varying levels of domain, strategic, and metacognitive knowledge that impact the plans they create. The *enactment* phase involves the interplay of many of the variables from the anticipatory stage as the learner works through task. Here, metacognitive processes oversee performance by utilizing monitoring judgments related to comprehension and strategy efficiency and subsequently control processes that regulate strategy toggling and adaptation. In addition, at the enactment stage the learner seeks out the learning environment and sources of help that are most beneficial for their performance. Working memory demands are also a key variable at this stage of the task that can determine the efficiency and quality with which one can ultimately process information. The *self-reflection* phase is a chance for the learner to tune and restructure goals, beliefs, strategies, interest and knowledge integration. This phase is key not only as an updating process for subsequent revolutions of task engagement beginning with the anticipatory stage but also one's SRL at the domain and dispositional level which are described below.

A learner's ability to develop SRL skills at the task level is also determined by others within the learning context, which might include same age/ability peers, older/higher ability peers, parents, tutors, or teachers. Interactions with these individuals might be either SRL-supportive or SRL-restrictive depending upon how they encourage intrinsic motivation, exploration of various cognitive strategies, choices, and the extent to which they model and scaffold monitoring processes, debugging strategies, and reflection. The medium of the learning session also plays a significant role in the task-level development of SRL. The medium may take many forms such as independent text-based sessions, direct instruction lessons in a classroom, inquiry-based classroom lessons, or computer-based sessions that may be either static, interactive, or immersive in nature. In addition, various incentives may impact motivation, help seeking, and other processing stages at the task level.

While most current SRL models have focused on representing the learner's cognitive, metacognitive, and motivational process at the task level, we speculate on an extension of those models that includes two broader levels in the SRL process. There is a vast literature illustrating that learners vary significantly in their level of expertise, metacognitive skills, level of self-efficacy, and goals adopted by domain (Chi, Glaser, & Farr, 1988; Ericsson, Charness, Feltovich, & Hoffman, 2006). Thus, it is reasonable to expect differences between and within individuals with regard to SRL by domain due to varying levels of expertise. Differences in working memory efficiency, automaticity of problem solving, and sophistication of knowledge-based schemas could impact a learner's ability to self-regulate effectively, particularly as the nature of the problem becomes more difficult. The model in Figure 3 also shows that over time experiences across domains inform more general skills that begin to resemble dispositions. At this level more enduring characteristics of the individual emerge, such as long-standing goals, beliefs and outlooks on the process of learning itself, and general metacognitive knowledge and skills that cut across domains. Like that which is seen at the domain level, empirical evidence exists to support SRL skills at the disposition level. For instance, monitoring accuracy has been shown to have a domain-general influence and to cluster within broad abilities such as fluid and crystallized ability (Schraw, Dunkle, & Bendixen, 1995; Schraw & Nietfeld, 1998). As a caveat, it is important to point out that the nature of this multi-level SRL model is speculative. Surely, there are more variables and detail that could be included at all three levels. Unfortunately, a justified explanation of the model and all of its component parts is beyond the scope of this chapter.

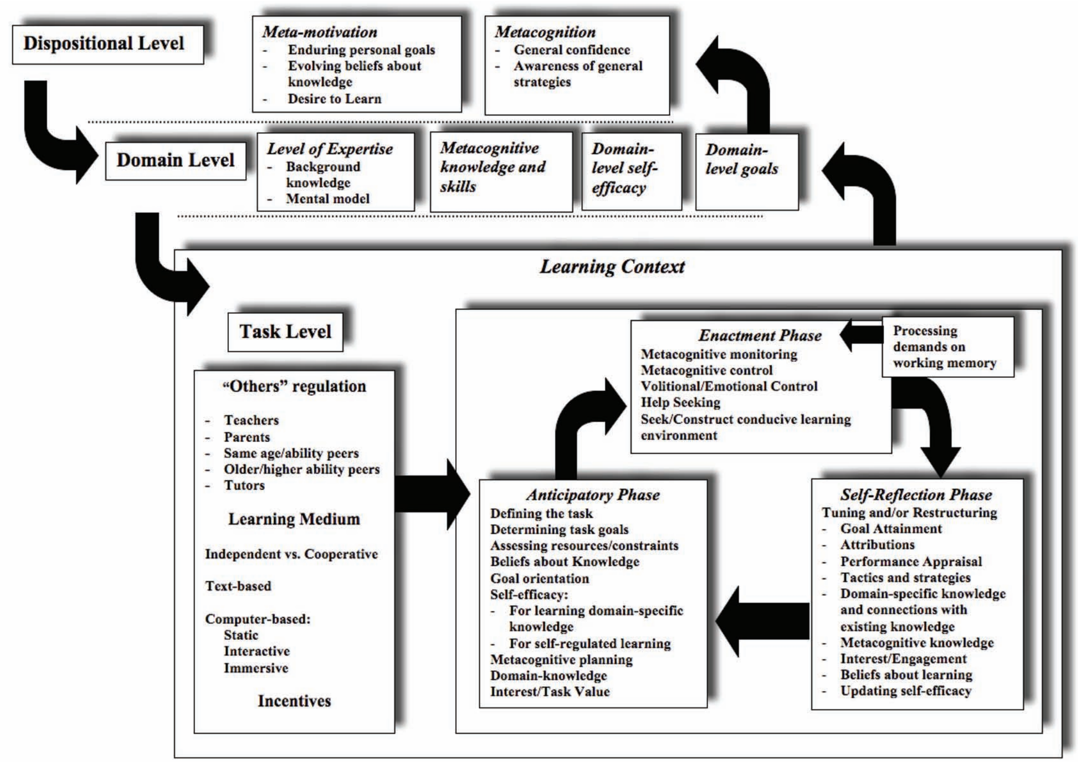


Figure 3. Multi-level model of self-regulated learning. The task level was informed by Winne, P. H., & Hadwin, A. E. (1998); Zimmerman (2000); and Pintrich (2000).

RECENT RESEARCH THAT INFORMS SELF-REGULATED LEARNING IN GAME-BASED LEARNING ENVIRONMENTS

SRL in Traditional Learning Environments

The bulk of existing SRL research has been conducted in traditional learning environments both within the context of on-going classroom instruction and lab-based studies. Overall, the findings are optimistic. Zimmerman and Tsikalas (2005) have concluded that, "Research on self-regulation in traditional learning environments has shown that sequential multilevel training not only enhances the development of metacognitive competence, but also enhances students' motivational self-beliefs, such as self-efficacy, which can sustain proactive learning even in the face of daunting obstacles" (p. 71).

Measurement presents one of the most formidable challenges for research in SRL. It is evident from previous research that effective learners exhibit strong SRL qualities; however it is somewhat of a leap to assume that all learners can and/or will self-report these qualities accurately. Although several self-report measures have been experimentally and empirically validated, their ecological validity for predicting course performance is questionable given this limitation of accurate self-assessment (Winne & Jameson-Noel, 2003). Similarly, Schunk (2008) contends that data is lacking on the extent to which SRL self-report measures accurately reflect actual behaviors that they were intended to measure. Examining self-regulatory constructs and their relationships inevitably involves the validity of their measures, which determines the quality of the research and impacts educational practice. Given that validity rests in the interpretation and use of scores within context (Messick, 1987), one persistent challenge is to provide evidence-based recommendations on the use of particular SRL scales and variables.

A full discussion of the relevant findings in SRL is too broad a task for this chapter (see Boekarts, Pintrich, & Zeidner, 2000; Schunk & Zimmerman, 2008 for more comprehensive reviews), thus we provide a brief summary below of some major findings organized into sections addressing strategy use, metacognition, and motivation.

Strategy use. VanZile-Tamsen and Livingston (1999) found that high- and low-achieving students differed on self-regulated strategy use. High achievers reported a significantly greater degree of engagement in such strategy use than did the low-achieving students. Unfortunately, the strategy reported being used most frequently by the majority of undergraduate students, rereading the textbook (Cao & Nietfeld, 2007), is considered a relatively ineffective approach to learning, as it is not active and involves surface level processing (Craik & Tulving, 1975). In contrast, a positive correlation has been found between active strategies and text comprehension (Pressley & Afflerbach, 1995). Good strategy users have a broad repertoire of strategies, metacognitive knowledge about those strategies, and are able to block out interference when employing strategies (Pressley, Borkowski, & Schneider, 1987).

Metacognition. Metacognitive monitoring skills and the regulation of strategies and tactics are core components within information processing models of self-regulation (Butler & Winne, 1995). More accurate monitoring has been shown to lead to improved self-regulation that, in turn, translates into improved performance (Thiede, Anderson, & Therriault, 2003). While accurate monitoring by students is certainly not a given (Glenberg & Epstein, 1985; Pressley, Ghatala, Woloshyn, & Pirie, 1990), metacognitive abilities in general are considered to be malleable and largely independent of general intelligence (Pressley & Ghatala, 1990). Students with effective metacognitive skills accurately estimate their knowledge and develop effective plans for new learning (Lan, 2005; McCormick, 2003; Thiede, et al., 2003). There is also evidence that monitoring accuracy on test performance can be improved with training and feedback for undergraduates (Nietfeld, Cao, & Osborne, 2006) and with elementary students who undergo comprehension monitoring training (Huff & Nietfeld, 2009). However, undergraduates with simple repetition of judgments without training or feedback have not shown gains in monitoring accuracy (Nietfeld, Cao, & Osborne, 2005). The lack of calibration in undergraduates carries over into study preparation in the form of overconfidence in their use of study tactics (Winne & Jamieson-Noel, 2002). In order to more fully understand how to improve SRL skills there is a great need for further research in how to improve both monitoring and control processes. In addition, these studies need to be conducted in externally valid, naturalistic settings (Hacker, Bol, & Keener).

Motivational control. There is a broad consensus that the development and utilization of metacognitive knowledge and metacognitive skills is mediated by motivational factors (e.g., Boekaerts, 1995, 1997; Schunk, 1995, Winne, 1995; Zimmerman, 2000). More specifically, the quality of the self-regulatory skills that students employ depends in part on several underlying beliefs that students hold about themselves. Key among these is self-efficacy beliefs: students' judgments of their capability to accomplish tasks and succeed in activities (Bandura, 1997). Self-efficacy beliefs predict academic performance (Schunk & Swartz, 1993; Schunk & Zimmerman, 2003) and function as a mediating variable that influences students' motivation to engage and sustain self-regulatory efforts (Bandura, 1997; Pajares, 1997). Students with high self-efficacy engage in more effective self-regulatory strategies, such as monitoring their academic work time effectively (Pintrich & De Groot, 1990) and are more confident in setting academic goals (Zimmerman & Bandura, 1994).

Implicit theories of intelligence and goal orientation also contribute to SRL (Dweck & Leggett, 1988; Ames & Archer, 1988; Elliot & McGregor, 2001). Individuals who adopt mastery goals have been shown to elicit numerous positive behaviors related to academic engagement (see Bruning, Schraw, Norby, & Ronning, 2004 for a more complete review). For instance, Schunk (1996) found that 4th graders who worked towards a learning goal had higher motivation and achievement outcomes than their 4th grade counterparts who worked toward performance goals. Learner interest in the task is another motivational variable that impacts SRL and can affect engagement and attention (Hidi, 1990) as well as depth of processing (Schiefele, 1999, Schraw, 1997). Interest can be distinguished between personal and situational interest (Schraw & Lehman, 2001). Personal interest is enduring and context-general whereas situational interest is spontaneous and context-specific. Capturing interest through one of these two forms can be critical for maintaining sustained engagement.

Despite the extensive research on cognitive and motivational constructs, there is a general agreement that the relationships among self-regulatory constructs are largely unknown and that future research should continue to address relationships among self-regulatory measures in order to advance research and educational practice (Schraw & Impara, 2000; Sperling, Howard, Staley, & Dubois, 2004). Self-regulation is developed through an intentional, active process facilitated through interventions (Schunk & Zimmerman, 2003). For instance, interventions for elementary school children that have focused on process goals plus feedback have been successful in traditional classrooms (Schunk & Swartz, 1993).

SRL Research in Computer-based Learning Environments

In one of the most comprehensive reviews of studies involving SRL variables with CBLEs, Winters, Greene, and Costich (2008) reviewed 33 empirical studies. Some of their findings applicable to GBLEs include:

- Students who utilize SRL processes more frequently show learning gains.
- Students with high prior knowledge utilize more SRL strategies.
- Mastery goal orientation is related to an increase in SRL strategies.
- Students with low SRL skills tend to excel more in program-controlled conditions while high SRL learners perform better in user-controlled conditions.
- On collaborative tasks, students regulate each other’s learning by maintaining mutual understanding and using cognitive strategies, but little monitoring or planning occurs.

Winters, et al. (2008) went on to report that students generally viewed built-in SRL tools as effective, however they failed to use these tools consistently. In addition, the authors reported that SRL skills are trainable but that students show little evidence of effective monitoring ability without such training. Recommendations from this review for future research included determining effective tactics for fading scaffolds with CBLEs, addressing individual differences in SRL, and a stronger focus on the quality rather than quantity of the SRL tasks.

Zimmerman and Tsikalas (2005) stress that in order to be effective at producing self-regulated learners, self-regulatory processes must be supported during the forethought, performance, and self-reflective phases of learning. In their review they outline how certain CBLEs support a subset of these phases, but rarely all three. In other words, certain systems (in particular, *Inquiry Island, Digital IdeaKeeper, iStart, Artemis, Cognitive Tutor,* and *AutoTutor*) are equipped with spaces for goal-setting and planning, monitoring, question-generation and collaboration, and self-evaluation and adaptation; however, no one system provides support during all three SRL phases. Zimmerman and Tsikalas (2005) stress the importance of providing instruction about how, when, and why to enact certain SRL processes and accomplishing this through instruction that includes three levels: an observation level, an emulation level, and finally a self-directed level. Zimmerman and Tsikalas note that both *iStart* (McNamara, Levinstein, & Boonthum, 2004) and *Cognitive Tutor* (Mathan & Koedinger, 2005) follow these three phases. First, expert models execute certain tasks and the learner simply observes. Second, the students attempt to emulate the expert models' processes and receive feedback on their responses. However, neither of these systems fade support of these skills after appropriate emulation success is achieved.

Intelligent tutoring system environments. Intelligent tutoring systems are computer-based simulations of traditional human-to-human tutoring sessions. Generally, the design of ITS environments are informed by the behaviors and techniques expert tutors convey during tutoring sessions and their ability to adapt to the student's current understanding and learning speed. Empirical evaluations of intelligent tutoring systems have shown significant learning gains for content, but significantly less attention has been paid to SRL (Winters, et al., 2008).

An exception to this is the work by Biswas and colleagues (Biswas et al., 2004, Biswas, Roscoe, Jeong, & Sulcer, 2009) using a teachable agent ITS environment called *Betty's Brain*, in which the computer agent is the "student" and the student takes on the role of the tutor (Biswas et al., 2004). *Betty's Brain* was created for teaching middle school science by developing a system that reasons about and forms conclusions based upon student-generated concept maps. Empirical investigations that have examined versions of *Betty's Brain* with and without integrated SRL strategies show students are better equipped for learning novel concepts after interacting with the SRL version of the system that explicitly prompts use of SRL strategies (Biswas, et al., 2009).

A study by Clarebout & Elen (2006) revealed somewhat surprising results when they found advantages for fixed time interval scaffolding, as opposed to adaptive scaffolding, from a pedagogical agent in an open-ended learning environment problem with undergraduates. The authors concluded that little is known about the efficacy of adaptive scaffolding in open learning environments and argue that most studies employing pedagogical agents have taken place in more constrained environments that focus on domain-specific knowledge. Their findings suggest a greater focus be placed on metacognitive abilities and self-regulation in open-ended learning environments where more choices are available to the learner. In our work with an open-ended GBLE named CRYSTAL ISLAND (to be described more fully below) we have attempted to employ prompted conditions that support SRL tactics. One initial study of such prompts however revealed that prompted conditions did not show any advantages over non-prompt conditions (Nietfeld, Shores, & Hoffmann, 2010). This finding could be due to the nature of the prompts themselves or likely that the prompts added cognitive load for students that would not reveal payoffs until automaticity was achieved in their implementation (Carlson, Chandler, & Sweller, 2003). This would require interaction with CRYSTAL ISLAND in a repeated sessions approach; a goal for future research. Thus, future research in open-ended environments whether they be game-based or not much consider multiple factors when constructing SRL-based scaffolds.

In their study of scaffolds for help-seeking, Roll, Aleven, McLaren, and Koedinger (2007) noted that while there are an increasing number of educational technology systems being developed to support metacognition few actually *teach* students the skills to become better future learners. Rather, more frequently the support provided by these systems function as a *crutch* that may produce short-term learning benefits rather than as a *scaffold* to help students internalize means of effective learning. Moreover, students must be motivated to use the metacognitive aids provided to them within tutoring systems. Koedinger and Aleven (2007) offered a number of recommendations for feedback in ITS systems utilizing cognitive tutors. These included the importance of providing immediate feedback, including goal feedback after errors on the part of the learner, an emphasis on mastery instruction that discourages “gaming,” and having learners provide self-explanations after receiving feedback for both their correct and incorrect responses.

Hypermedia and strategy use. Hypermedia environments present students with a rich series of multimedia pages and allow the student to freely navigate through the environment with the use of hyperlinks in order to achieve the learning goal. Since hypermedia environments can be quite large, students are required to utilize SRL skills in order to deduce what specific information is necessary, how to locate that information within the environment, and most importantly how to integrate that information and fully understand the learning goal. That being said, the use of the SRL strategies summarization and knowledge elaboration has been shown to yield qualitative mental model shifts in these environments (Greene & Azevedo, 2005). In addition, Moos and Azevedo (2009) found specific monitoring processes to be related to both self-efficacy and prior knowledge for undergraduates learning in a hypermedia learning activity.

Help seeking, a student to teacher collaboration method, is a commonly emphasized in models of SRL (Pintrich, 2000). Azevedo, Cromley, Winters, Moos, and Greene (2005) found that adaptive scaffolding in a hypermedia environment increases the likelihood of students engaging in help seeking activities. The authors suggest that help seeking indicates a student’s desire to learn about an advanced topic and therefore, this activity must be scaffolded in order for students to master complex topics. Therefore, the authors suggest creating databases to anticipate questions students might pose and to pair these questions with corresponding scaffolds. Winters & Azevedo (2005) found that prior knowledge impacted help seeking in CBLEs. They investigated the collaboration behaviors of low and high prior knowledge high school aged dyads during interactions with a CBLE on genetics and found the following: low prior knowledge students rely on their partners for regulatory support more than the opposite, high prior knowledge students set self-created goals rather than using the teacher-formulated goals, and low prior knowledge students often sought help from their high prior knowledge partner for clarification of confusing material.

Research on SRL within GBLEs

Existing research in GBLEs specifically targeting SRL variables has been very limited. In the 19 peer-reviewed empirical studies reviewed by O’Neil, Wainess, and Baker (2005) regarding the effectiveness of digital games, none categorically addressed SRL. Some still beg the question of whether computer games really foster an engaging, effective learning experience in classrooms (Ke, 2008), but with the current research momentum in GBLEs it seems more appropriate to ask *how* can we *build* games that foster learning *and* learning how to learn. O'Neil et al (2005) and Garris, Ahlers, & Driskell (2002) reiterate this sentiment by noting that games themselves are not sufficient for learning but elements within games can be activated to encourage learning. Thus, instructional support is essential. The consensus thus far appears to be that there is a lack of sound instructional design embedded in GBLEs (O'Neil et al, 2005).

Most GBLE studies have reported findings related to content knowledge. Papastergiou (2009) found a game-based learning approach was more effective in increasing students' knowledge of computer memory concepts and was more motivational than for students using a non-gaming approach on computer. Barab et al., (2007) showed learning gains and transfer on a distal-level multiple-choice assessment after a 3-week curriculum within *Quest Atlantis* involving an ecology-based problem with 5th grade students functioning as field researchers in the game.

One GBLE that is being evaluated for its ability to encourage SRL-related variables in addition to its primary focus as a problem-based learning environment is *Alien Rescue*. *Alien Rescue* is a game for sixth-grade students that is intended to teach students about astronomy and space travel over repeated sessions. Hseih, Cho, Liu, & Schallert (2008) had students play *Alien Rescue* everyday for 3 weeks in their science class in dyads or groups of three. After the intervention, students' science achievement scores and self-efficacy levels for science increased significantly even in the absence of direct instruction. Moreover, both performance-approach and performance-avoid goal orientations decreased after interacting with Alien Rescue. A slight increase in mastery-goal orientation was found. Liu, Hsieh, Cho, & Schallert, (2006) also found increases in self-efficacy after students played Alien Rescue over repeated sessions and found self-efficacy to be predictive of science achievement.

Ke (2008) examined a math drill and practice game for a small group of 4th and 5th grade students and found no changes in knowledge or on a measure of meta-cognitive awareness (jrMAI) but did find increased attitudes towards math during 10 two-hour sessions. The lack of change in metacognitive awareness however, is not surprising given that there was not an intervention built within the game to impact metacognition.

Narrative-centered learning environments (NLEs) have received attention for their potential benefits on student motivation. NLEs are being developed in many domains including language learning (Johnson, Vihjalmsson, & Marsella, 2005), anti-bullying education (Aylett, Louchart, Dias, Paiva, & Vala, 2005), and middle school science (Ketelhut, 2007; McQuiggan, Rowe, Lee, & Lester, 2008). NLEs involve rich, immersive storyworlds through which learning occurs as the student, the protagonist of the narrative, experiences and unfolds the narrative. Generally NLEs employ structural characteristics, such as novelty and intensity, as well as content characteristics, such as human activity and life themes, all of which have been shown to contribute to situational interest (Ainley, Hidi, & Berndorff, 2002; Hidi, 1990).

Fantasy contexts in educational games have also been shown to provide motivational benefits for learning (Parker and Lepper, 1992). Narrative features such as pacing and tension can introduce additional challenge to learning tasks and contribute to student motivation. Ultimately, NLEs rely on the inherent structure of narratives and student contributions to guide the progression of the plot; therefore, the narrative is devised in such a way that scaffolds the student to gain an understanding of the presented material and integrate those understandings with problem-solving techniques to achieve the desired resolution (Barab, et al. 2009). Involving students in GBLEs requires an emphasis on situative embodiment wherein the learner engages in a narratively-rich setting, having goals, a legitimate role, and actions that have consequences associated with them (Barab, et al., 2007). Being in the context and solving problems related to the context is essential. Barab et al., (2007) argue that well-designed game play embodies players perceptually, socially, and narratively. Moreover, they argue that K-12 curriculum would be more "useful" if the content was similarly embodied in interactive narrative contexts.

Immersing the learner within a narrative by assigning a particular role to the student's character allows for the potential to discreetly embed several SRL techniques to maintain self-efficacy, increase calibration, and scaffold strategy use (for a more detailed discussion see Shores, Robison, Rowe, Hoffmann, & Lester, 2009). Also, since NLEs can be programmed for multi-session game play, the duration of SRL training can be extended in order to increase the likelihood that students will internalize self-regulatory skills. Not only do NLEs allow for noninvasive SRL instruction, but also SRL metrics. Note-taking, expert character help seeking, on-task behavior, and goal setting and monitoring can easily be logged and quantitatively analyzed.

Incorporating SRL approaches from traditional environments to NLEs is not without its challenges. Our work has involved assisting in the development an inquiry-based NLE called CRYSTAL ISLAND (see Figure 3) in the domain of microbiology for 8th grade students. CRYSTAL ISLAND features a science mystery set on a recently discovered volcanic island where a research station has been established to study the unique flora and fauna. The student plays the protagonist attempting to ultimately discover that the milk on the island is carrying an unidentified infectious disease by utilizing resources at the research station. The story opens by introducing the student to the island and the members of the research team. As members of the research team fall ill, it is the student's task to discover the cause of the specific source of the outbreak. This involves exploring the world and interacting with other characters while forming questions, generating hypotheses, collecting data, and testing the hypotheses. In one version of the mystery there is a poisoning scenario involving one of the research team members that adds to the narrative. Throughout the mystery, the student can walk around the island and visit the infirmary, the laboratory, the dining hall, and the living quarters of each member of the team. The student can pick up and manipulate objects, talk with characters to gather clues about the source of the disease. Facts and clues revealed during the student's interaction can be stored in a virtual diagnosis worksheet. Some characters within the game serve as local experts on certain pathogenic diseases. During the student's interactions with

Figure 3. Crystal Island.

such expert characters, in-game quizzes are presented to test the student’s knowledge and to give them the opportunity to gain more privileges for testing objects on the island. In the course of the adventure, the student must gather enough evidence to correctly identify that the milk has been contaminated with E-Coli. To win the game, the student must submit a correct diagnosis worksheet with information about the source object, disease, and treatment to the camp nurse for review.

Interaction with CRYSTAL ISLAND has produced significant gains in content knowledge and has also revealed higher levels of reported presence (feeling of being immersed in the game) when a full narrative versus a minimal narrative condition was employed (McQuiggan, et al., 2008). Nietfeld, et al. (2010) found that the ability to correctly identify components of the science mystery on the diagnosis worksheet and the ability to answer content-related questions during play in CRYSTAL ISLAND predicted application-level performance even when controlling for pretest differences in prior knowledge. These findings added support for CRYSTAL ISLAND as an inquiry-based environment that encourages problem-solving skills and facilitates higher-level knowledge.

Studies with CRYSTAL ISLAND have also begun to provide some illuminating results regarding SRL. In one study students provided measures of goal-orientation, monitoring, and situational interest before and after their participation with CRYSTAL ISLAND (Nietfeld, Hoffmann, McQuiggan, & Lester, 2008). Students were randomly assigned either a learning goal or a performance goal before playing and were also randomly assigned feedback after playing indicating that their performance was either high or low. Results indicated that goal orientation and situational interest were significantly affected by feedback related to their performance. For instance, students who were assigned a learning goal and who received a high scoreboard (good performance relative to peers) showed a significant decrease in performance goals and students who were assigned a performance goal and who received a low scoreboard (poor performance relative to peers) showed a decrease in performance approach goals. Students assigned to learning goal conditions showed greater interest after playing CRYSTAL ISLAND and students who had higher mastery approach scores going into the game showed higher interest ratings when given a learning goals as opposed to a performance goal. Additionally, monitoring judgments made throughout gameplay were a key predictor of performance measures such as goals completed ($r = .59$) and score ($r = .74$) within CRYSTAL ISLAND.

Other interesting findings related to motivational variables with CRYSTAL ISLAND have been found. In the McQuiggan et al. (2008) study students with higher science efficacy reported higher levels of presence than their less efficacious peers and mastery-oriented students reported higher levels of presence than performance-oriented students. In another study, interest ratings for CRYSTAL ISLAND were found to predict problem-solving transfer after controlling for prior microbiology knowledge, self-efficacy for science, and gaming interest (Shores, Hoffmann, & Nietfeld, 2010). Transfer was predicted by interest on each of four tasks varying in degrees of near to far transfer. Differences in motivation have also been found between genders in CRYSTAL ISLAND (Nietfeld, Hoffmann, & Shores, 2010). Eighth grade girls attributed their performance in CRYSTAL ISLAND as being due to luck and task difficulty more so than eighth grade boys but boys attributed their performance more to talent than girls. Differences were also found for goal orientations and interest. For instance, there was a significant negative relationship between performance approach goal orientation and goals completed ($r = -.42$) in CRYSTAL ISLAND for girls but not for boys. Boys, but not girls, who were more interested in CRYSTAL ISLAND tended to complete more goals ($r = .40$). These last findings show the potentially important role of adaptive scaffolding to suit individual differences, in this case gender, to provide a more customized approach in teaching SRL skills.

CHALLENGES AND FUTURE DIRECTIONS FOR MEASURING AND ENHANCING SRL IN GAME-BASED LEARNING ENVIRONMENTS

The research discussed above gives an indication of the high level of emphasis the research community is placing on developing SRL skills and GBLEs that improve learning. Moving forward, collaboration between those doing research on SRL in traditional settings, those in the ITS field, and those in the gaming world is essential. In addition, the accessibility of GBLEs for schools is of high importance. Researchers who create GBLEs to be used in the schools must consider the technologies available at the schools they will be serving, consider the role of the teacher, the curriculum and curriculum pacing, and the relationship of classroom instruction with the created GBLE. Below, we offer several suggestions for future research related to SRL in GBLEs and we organize these suggestions according to the levels within the model of SRL presented at the beginning of this chapter. GBLEs vary significantly and it is beyond the scope of this chapter to be able to customize the following suggestions for any given environment. Therefore, the suggestions are intended to be more general to the field and can be adapted for any particular GBLE. The following sections are organized in a top-down fashion according to our proposed model of SRL. Given the speculative nature of the dispositional- and domain-levels as well as the limited empirical data currently available, discussion will be limited relative to the task-level.

Dispositional-level Challenges

The efficacy of GBLEs to impact dispositional-level aspects of SRL is untested and a far-reaching challenge for future research. Can game-based environments actually influence relatively stable domain-general belief systems and orientations that a person holds? One potential starting point for this investigation might be for GBLEs to build in opportunities for students to develop formal theories about their cognitive and motivational processes as they reflect on their performance (Schraw & Moshman, 1995). Some researchers have already begun this by developing their software to make explicit to the learner aspects of metacognition such as task structures, the importance of planning and reflection, the space of activities that learners need to monitor, and the regulation aspect of strategies (Quintana, Zhang, & Krajcik, 2005). This type of approach is also one that actively encourages transfer of knowledge. Other approaches that GBLEs might take to encourage dispositional change is confronting naïve beliefs, including higher-level problem-solving activities that cut across domains, and emphasizing general strategies that can be used across games and settings. Dispositional-level confidence and self-efficacy might be impacted if students are given the opportunity in GBLEs to take risks and approach problems in ways that they might be reluctant to in real life. In addition, students could be asked to reflect on dispositional-level traits, such as personal goal orientation, and then be given comparison feedback regarding such traits after their GBLE interactions.

Domain-level Challenges

Schraw, Crippen, and Hartley (2006) point out that work specifically targeting the development of SRL skills within the domain of science has been limited. Given that this claim is accurate, the work involving SRL within science-based GBLEs has been even more limited and this is most likely common across domains. If GBLEs are to have an impact on SRL at the domain level research is needed examining GBLEs equipped with SRL scaffolding capabilities, a significant amount of curriculum coverage, and problem-solving scenarios complex enough to require students to interact with the system over a long duration of time. Flexible knowledge and skills are also more likely to develop if the content and/or targeted SRL skills are augmenting skills taught in the classroom or other educational context (Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987). Teaching students meaningful SRL skills will likely take months and vary depending upon developmental level and implicit theories of learning (Hofer, Yu, & Pintrich, 1998).

Task-level Challenges

The scepticism by those who doubt the efficacy of GBLEs to positively impact learning will likely be justified if such environments fail to teach students *how* to learn while they are gaming just like it is essential to teach students *how* to learn in a traditional classroom setting. This translates into functionalities that model and scaffold goal setting and tracking, presenting and employing effective cognitive strategies, using appropriate metacognitive monitoring and control processes while in the midst of problem solving, and regulating various motivational processes at all junctures of the activity. In sum, the recommendation for CBLEs by Zimmmerman and Tsikalas (2005) is to build in support for all 3 phases of the task level, anticipatory, enactment, and self-reflection, applies also to GBLEs.

Engagement and learning. One challenge at the task level for developers of serious games is creating an environment that is both fun for the students while at the same time a sophisticated learning environment. In other words, techniques to promote SRL must be as seamlessly integrated as possible. This is important given that there is evidence showing that engagement functions as a gateway to learning in GBLEs (Rowe, Shores, Mott, & Lester, 2010). In their study, Rowe et al. (2010) investigated the effect of student engagement on content learning gains in CRYSTAL ISLAND and revealed that student engagement as measured by presence, interest, and in-game score, all predicted performance on a content knowledge post-test while controlling for previous background knowledge. There are likely many instances where an SRL technique that is successful in traditional learning environments will have to be adapted within a GBLE. For instance, research in traditional settings for learning from text (Thiede, et al., 2003) and within classroom settings (Nietfeld, et al., 2006) has shown the importance that monitoring judgments can make on comprehension and overall performance. We have learned from our work with CRYSTAL ISLAND that simply inserting these types of prompts that require students to make accuracy and performance-related judgments disrupts the flow of the game and can be viewed by the student as irrelevant and out of context. An alternate approach, however, is to have students provide the same type of judgments but do so within the storyline as part of the narrative through communication with other characters in the environment. For instance, we are currently experimenting with our main character (Alyx) communicating with other members of the research team on the island through the use of a personal digital assistant in order to report monitoring judgments couched within the narrative dialogue. Another example of implicit integration of SRL within the narrative is accomplished by assigning goals through the character chosen by or assigned to the student (White & Frederiksen, 2005). This is done in *Inquiry Island* in an attempt to have elementary students internalize metacognitive skills through the character roles that they adopt. Future research will determine creative ways of integrating other SRL features such as maintaining persistence and self-efficacy, strategy toggling, and appropriate help seeking.

Measurement of SRL variables. Just as the measurement of SRL is a challenge in traditional learning environments it will remain one of the most formidable challenges for SRL research within GBLEs. Creative means of assessing learners both implicitly, such as through trace data (Hadwin, Nesbit, Code, Jamieson-Noel, & Winne, 2007) and stealth assessment (Shute et al., 2010) during game play and explicitly to provide students with the opportunity to actively monitor and reflect on their learning will be needed. Trace data can be used to test or verify other forms of assessment such as self-report measures and multiple-choice items. Winne and Nesbit (2009) provide some helpful data tracking guidelines for SRL and metacognition wherein they recommend collecting physiological measures, performance measures, on-line self-report probes, distal self-report probes, eye-tracking data, and software logs. In addition to assessment related directly to the GBLE experience itself, it will be essential to develop rich, performance-based assessments to measure the transfer of skills outside of the GBLE. These are often more difficult to create and validate but can provide a more meaningful measure of the effectiveness of the GBLE. Irrespective of how well these recommendations are followed, the contextual nature of GBLEs will continually challenge researchers in the measurement of multi-faceted abilities such as metacognition in a manner that is non-invasive to students' game play.

Adaptive scaffolding. Advances in adaptive scaffolding approaches in GBLEs will be extremely important to accommodate for individual differences (gender, low prior knowledge, gaming experience, etc.) or developmental levels particularly in open-ended and less structured environments (Kirschner, Sweller, & Clark, 2006). Predictive modeling techniques such as Bayesian approaches have shown some success thus far in predicting learner metacognition and motivation (McQuiggan, Hoffmann, Nietfeld, Robison, & Lester, 2008; McQuiggan, Mott, & Lester, 2008). However, these approaches utilize probabilities from actions in trace data and will become more tenuous for environments that contain an academic year's worth of curriculum where students spend hours or weeks navigating the environment. More studies are also needed on the varying forms of scaffolding such as hints from pedagogical agents, highlighted locations/objects in the world (assuming an immersive environment), changes in the world or structure of the activity, or user controlled scaffolds. As in traditional learning environments, many SRL skills such as monitoring and control processes, will likely not show improvement without training, therefore creative ways of integrating this training will have to be discovered.

Physiological measures. Studies capturing physiological data such as eye-tracking have the potential to provide significant advances in understanding cognitive processing and motivational states in GBLEs. For instance, issues of cognitive load during the enactment phase of a task may hinder a user's strategy efficiency. Eye tracking devices have the potential to identify visual fixations in such contexts that could inform adaptive scaffolding capabilities. In addition, such devices could assist in determining how learners refer to various forms of text, particularly when voice and text are interacting within the same environment. This data could inform developers as to whether, for instance, a toggling device should be provided to the user as an option to go back and forth between text and voice. These types of functions could also benefit second language learners in their attempt to self-regulate their study under greater processing challenges. Merton and Conati (2006) have successfully used eye-tracking data to assess metacognitive behavior (self-explanation) in an open-ended learning environment and subsequently improve on modeling capabilities that previously used only student interface actions. The more precise student models can then be used to generate a more adaptive experience for the learner. D'Mello and Graesser (2009) have utilized other physiological data, namely posture patterns via seat sensors, in addition to eye-tracking data to accurately detect student affect while interacting with *AutoTutor,* an ITS.

Collaboration. The study of collaboration within GBLEs and its impact upon SRL will be another area of interest at the task level. In non-gaming environments investigation is underway examining how the use of various technologies for collaboration amongst students impacts SRL (Morris et al, In Press). In GBLEs collaboration has been found to improve metacognitive abilities in elementary children in *Inquiry Island* (White & Frederiksen, 2005). GBLE studies of collaboration can be extended by investigating user chats, multi-player games both virtual and side-by-side, and with companion agents. In addition, collaboration can come in the form of teacher and student interaction with teacher-developed games (Annetta, 2008).

Self-explanation. One means of encouraging and assessing SRL in GBLEs is to have students generate self-explanations for their problem-solving processes. Aleven and Koedinger (2002) found that high school students who utilized a self-explanation strategy through a geometry Cognitive Tutor showed higher transfer scores and more well-integrate knowledge. A major goal in the assessment of SRL in GBLEs is to have students come to an explicit understanding of effective problem-solving processes. In order to do this they must first recognize and externalize their performance (monitoring) and then develop strategies to address their shortcomings (control).

CONCLUSIONS

In this chapter we have attempted to provide an overview of research that has been accomplished related to SRL and GBLEs and to provide some suggestions for future research. Although research in SRL and GBLEs are still in their infancy it is exciting to consider the potential that exists at the intersection of the two fields. In this chapter we have also put forth a tentative model of SRL from which to form a general framework for investigations within both traditional settings and GBLEs. Hopefully, the findings and remaining challenges reviewed here will be the basis of a jumping off point for future examinations of SRL with GBLEs.

REFERENCES

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*, 545–561.

Aleven, V., & Koedinger, K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26,* 147–179.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Student learning strategies and motivation processes. *Journal of Educational Psychology*, *80*, 260–267.

Annetta, L. A. (2008). Video games in education: Why they should be used and how they are being used. *Theory Into Practice*, *47*, 229–239.

Aylett, R., Louchart, S., Dias, J., Paiva, A., & Vala, M. (2005). *FearNot! An experiment in emergent narrative*. 5th International Conference on Intelligent Virtual Agents.

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, *33*, 381–412.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research*. New York: Longmans.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.

Barab, S., Zuiker, S., Warren, S., Hickey, D., Ingram, A., Kwon, E., et al. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education*, *91*, 750–792.

Barab, S. A., Dodge, T., Tuzun, H., et al. (2007). The Quest Atlantis project: A socially-responsive play space for learning. In B. E. Shelton & D. Wiley (Eds.), *The educational design and use of simulation computer games* (pp. 159–186). Rotterdam, The Netherlands: Sense Publishers.

Barab, et al. (2009). Transformational plays as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology*.

Biswas, G., Leelawong, K., Belynne, K., Viswanath, K., Vye, N., Schwartz, D., et al. (2004). *Incorporating self regulated learning techniques into learning by teaching environments*. Twenty Sixth Annual Meeting of the Cognitive Science Society, Lawrence Erlbaum Associates.

Biswas, G., Roscoe, R., Jeong, H., & Sulcer, B. (2009). Promoting self-regulated learning skills in agent-based learning environments. In *Proceedings of the 17th international conference on computers in education*. Hong Kong: Asia-Pacific Society for Computers in Education.

Boekaerts, M. (1995). Self-regulated learning: Bridging the gap between metacognitive and metamotivation theories. *Educational Psychologist*, *30*, 195–200.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*, 161–186.

Boekarts, M., Pintrich, P. R., & Zeidner, M. (Eds.), (2000). *Handbook of self-regulation*. San Diego, CA: Academic.

Bruning, R., Schraw, G., Norby, M., & Ronning, R. (2004). *Cognitive psychology and instruction* (4th ed.). Upper Saddle River, NJ: Merrill.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*, 245–281.

Cao, L., & Nietfeld, J. L. (2007). College students' metacognitive awareness of difficulties in learning the class content does not automatically lead to adjustment of study strategies. *Australian Journal of Educational and Developmental Psychology*, *7*, 31–46.

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology*, *95*, 629–640.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Clarebout, G., & Elen, J. (2006). Open learning environments and the impact of a pedagogical agent. *Journal of Educational Computing Research*, *35*, 211–226.

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

D'Mello, S. K., & Graesser, A. C. (2009). Automatic detection of learners' emotions from gross body language. *Applied Artificial Intelligence*, *23*, 123–150.

Dweck, C., & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273.

Elliot, A., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*, 501–519.

Ellis, D., & Zimmerman, B. J. (2001). Enhancing self-monitoring during self-regulated learning of speech. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 205–228). Dordrecht, The Netherlands: Kluwer.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441–467.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning*, *2*, 5–16.

Glenberg, A., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 702–718.

Graesser, A. C., McNamara, D. S., & Van Lehn, K. (2005). Scaffolding deep comprehension strategies through point & query, AutoTutor, and iSTART. *Educational Psychologist, 40*, 225–234.

Greene, J. A., & Azevedo, R. (2005). *Adolescent's use of SRL behaviors and their relation to qualitative mental model shifts while using hypermedia*. Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology, IOS Press.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170.

Hacker, D. J., Bol, L., & Keener, M. C. (2009). Metacognition in education: A focus on calibration. In J. Dunlosky, & R. Bjork (Eds.), *Handbook of memory and metacognition*. Mahwah, NJ: Erlbaum.

Hadwin, A. F., Nesbit, J. C., Code, J., Jamieson-Noel, D. L., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*, 107–124.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*, 549–571.

Hofer, B. K., Yu, S. L., & Pintrich, P. R. (1998). Teaching college students to be self-regulated learners. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 57–85). New York: Guilford.

Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning, 4*, 161–176.

Hsieh, P. H., Cho. Y., Liu, M., & Schallert, D. L. (2008). Examining the interplay between middle school students' achievement goals and self-efficacy in a technology-enhanced learning environment. *American Secondary Education Journal*, *36*, 33–50.

Johnson, W. L., Vilhjálmsson, H. H., & Marsella, S. (2005). Serious games for language learning: How much game, how much AI? In *Artificial intelligence in education* (pp. 306–313). Amsterdam, The Netherlands: IOS Press.

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers and Education*, *51*, 1609–1620.

Ketelhut, D. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in river city, a multi-user virtual environment. *Journal of Science Education and Technology*, *16*, 99–111.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Koedinger, K., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239–264.

Lajoie, S. P., & Acevedo, R. (2006). Teaching and learning in technology-rich environments. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 803–821, 2nd ed.). Mahwah, NJ: Erlbaum.

Lan, W. (2005). Self-Monitoring and its relationship with educational level and task performance. *Educational Psychology*, *25*(1), 109–127.

Liu, M., Hsieh, P., Cho, Y. J., & Schallert, D. L. (2006). Middle school students' self-efficacy, attitudes, and achievement in a computer-enhanced problem-based learning environment. *Journal of Interactive Learning Research*, *17*, 225–242.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, *40*, 257–265.

McCormick, C. B. (2003). Metacognition and learning. In W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (pp. 79–102). New York: Wiley & Sons.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Insteractive strategy trainer for active reading and thinking. *Behavior Research Methods, Instruments, and Computers*, *36*, 222–233.

McQuiggan, S. W., Hoffmann, K. L., Nietfeld, J. L., Robison, J., & Lester, J. (2008, June). *Examining self-regulated learning in a narrative-centered learning environment: An inductive approach to modeling metacognitive monitoring*. To appear in Proceedings of the ITS'08 Workshop on Metacognition and Self-Regulated Learning in Educational Technologies, Montreal, Canada.

McQuiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction*, *18*, 81–123.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-Based learning: The impact of narrative on learning experiences and outcomes. In *Proceedings of the ninth international conference on intelligent tutoring systems* (pp. 530–539). Montreal, Canada.

Merton, C., & Conati, C. (2006). Eye-tracking to model and adapt to user meta-cognition in intelligent learning environments. In *Proceedings of the 11th international conference of user interfaces* (pp. 39–46). Sydney, Australia.

Messick, S. (1987). *Validity (RR-87-40)*. Princeton, NJ: Educational Testing Service.

Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning*, *4*, 197–216.

Morris, R., Hadwin, A. F., Gress, C. L. Z., Miller, M., Fior, M., Church, H., et al. (In Press). Designing roles, scripts, and prompts to support CSCL in gStudy. *Computers in Human Behavior*.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance and monitoring accuracy. *Metacognition and Learning*, *2*, 159–179.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education*, *74*, 7–28.

Nietfeld, J. L., Hoffmann, K. L., McQuiggan, S. W., & Lester, J. (2008, July). *Self-regulated learning in a narrative centered learning environment*. Vienna, Austria: Annual meeting of Ed-Media.

Nietfeld, J. L., Hoffmann, K. F., & Shores, L. R. (2010, January). *Gender differences in motivation within a narrative-centered learning environment*. Annual meeting of Hawaii International Conference on Education, Honolulu, HI.

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research*, *95*, 131–142.

Nietfeld, J. L., Shores, L. R., & Hoffmann, K. F. (2010). *The impact of a narrative-centered learning environment on application-level science knowledge*. Unpublished manuscript.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal*, *16*, 455–474.

Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 1–49). Greenwich, CT: JAI Press.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*, 1–12.

Parker, L. E., & Lepper, M. R. (1992). The effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, *62*, 625–633.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaert, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulated learning* (pp. 13–39). San Diego, CA: Academic Press.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33–40.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillsdale, NJ: Erlbaum.

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategies users coordinate metacognition and knowledge. In R. Vasta, & G. Whitehurst (Eds.), *Annals of child development* (Vol. 5, pp. 89–129). Greenwich, CT: JAI Press.

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, *25*, 19–34.

Pressley, M., Ghatala, E., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly*, *25*, 233–249.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*, 235–244.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, *2*, 125–140.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (submitted for 2010, June). Integrating learning and engagement in narrative-centered learning environments. If accepted: *To appear in Proceedings of the Tenth International Conference on Intelligent Tutoring Systems (ITS).* Pittsburgh, PA.

Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, *3*, 257–280.

Schraw, G. (2007). The use of computer-based environments for understanding and improving self-regulation. *Metacognition and Learning*, *2*, 169–176.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36*, 111–139.

Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, *87*, 433–444.

Schraw, G., & Impara, J. C. (2000). *Issues in the measurement of metacognition.* Lincoln, NE: Buros Institute of Mental Measurement and Department of Educational Psychology, University of Nebraska-Lincoln.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review*, *13*, 23–52.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*, 351–371.

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, *90*, 236–248.

Schunk, D. H. (1995). Inherent details of self-regulated learning include student perceptions. *Educational Psychologist*, *30*, 213–216.

Schunk, D. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review*, *20*, 463–467.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing instruction. *Contemporary Educational Psychology*, *18*, 337–354.

Schunk, D. H., & Zimmerman, B. J. (2003). Self-regulation and learning. In W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (pp. 59–78). New York: Wiley & Sons.

Schunk, D., & Zimmerman, B. (2008). *Motivation and self-regulated learning: Theory, research, and applications*. New York: Erlbaum.

Shores, L. R., Hoffmann, K. F., & Nietfeld, J. L. (2010, February). *The impact of narrative-centered learning environments on situational interest and problem-solving transfer.* Savannah, GA: Annual meeting of Eastern Educational Research Association.

Shores, L. R., Robison, J. L., Rowe, J. P., Hoffmann, K. F., & Lester, J. C. (2009). *Narrative-centered learning environments: A self-regulated learning perspective*. The 2009 AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems, Washington, DC.

Shute, V. J., Masduki, I., Donmez, O., Kim, Y. J., Dennen, V. P., Jeong, A. C., et al. (2010). Modeling, assessing, and supporting key competencies within game environments. To appear in D. Ifenthaler,

P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer-Verlag.

Sperling, R. A., Howard, B. C., Staley, R., & Dubois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10*, 117–139.

Spiro, R. J., Vispoel, W. P., Schmitz, J. G., Samarapungavan, A., & Boerger, A. E. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex content domains. In B. K. Britton, & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 177–199). Hillsdale, NJ: Lawrence Erlbaum.

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*, 19–29.

Thiede, K. W., Anderson, M. C., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 66–73.

VanZile-Tamsen, C., & Livingston, J. (1999). The differential impact of motivation on the self-regulated strategy use of high-and low-achieving college students. *Journal of College Student Development, 40*(1), 54–60.

Wery, J., & Nietfeld, J. L. (2010). Supporting self-regulated learning with exceptional children. *Teaching Exceptional Children*, *42*, 70–78.

White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, *40*, 211–223.

Winne, P. H. (1995). Self-regulation is ubiquitous but its forms vary with knowledge. *Educational Psychologist*, *30*, 223–228.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*, 551–572.

Winne, P. H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, *28*, 259–276.

Winne, P. H., & Nesbit, J. C. (2009, April). *Data for modeling self-regulated learning*. San Diego, CA: Annual meeting of American Educational Research Association.

Winters, F. I., & Azevedo, R. (2005). High school students' regulation of learning during computer-based science inquiry. *Journal of Educational Computing Research*, *33*, 189–215.

Winters, F., Greene, J., & Costich, C. (2008). Self-Regulation of learning within computer-based learning environments. *Educational Psychology Review*, *20*, 429–444.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaert, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulated learning* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self regulatory influences on writing course attainment. *American Educational Research Journal*, *31*, 845–862.

Zimmerman, B., Tsikalas, K. (2005). Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning. *Educational Psychologist*, *40*, 267–271.

John Nietfeld and Lucy R. Shores
North Carolina State University