A Foundational Research Study of Zeos

July 19, 2012

cultivating learning and positive change

www.magnoliaconsulting.org
Acknowledgements

I would like to thank the staff at Pearson, especially Ann Vilcheck and Kaili Barone, for their time and support of this work. In addition, I would like to recognize several Magnolia team members who made this work possible, including Dr. Stephanie Baird Wilkerson, Beth Peery and Candace Rowland.

The author,

Mary Styers, Ph.D.

Magnolia Consulting, LLC
5135 Blenheim Rd.
Charlottesville, VA 22902
(ph) 855.984.5540 (toll free)
http://www.magnoliaconsulting.org
Table of Contents

Introduction........................................................................................................................................... 1
Zeos aligns to Common Core State Standards...................................................................................... 2
Zeos provides progress monitoring for teachers..................................................................................... 3
Zeos assesses student mastery through use of proprietary mastery algorithms............................... 4
Zeos provides opportunities for distributed practice .............................................................................. 6
Zeos incorporates elements of cognitive load theory.............................................................................. 8
Zeos incorporates elements of game-based learning design ............................................................... 10
Zeos provides students with reinforcement and immediate feedback .............................................. 12
Zeos acknowledges the influence of peers on motivation ................................................................. 14
Conclusions ........................................................................................................................................... 15
References ............................................................................................................................................. 16
Notes ..................................................................................................................................................... 23
Introduction

Nationwide, educators and legislators are focusing on ways to improve educational accountability and student achievement. Initiatives such as No Child Left Behind (NCLB) (2002), new waivers to provide states with alternatives to NCLB standards (Duncan, 2012) and the recent development of Common Core State Standards (National Governors Association and the Council of Chief State School Officers, 2011) have led states and districts to seek academic programs that share a common purpose with these initiatives.

Pearson developed the online program Zeos to help students in grades 3–10 prepare for state and Common Core high-stakes assessments in English language arts and math. When students log in to Zeos for the first time, they create a superhero character and learn test-taking tips from other comic-based superheroes. Teachers or students can assign “challenges” or quizzes in the program. Students learn immediately if their answers are correct or incorrect, and receive additional feedback and support for incorrect answers. Once students have mastered the skills in the assigned challenge, they are able to spin a reward spinner to receive premium items for their superhero character. Students also are able to interact with peers on the platform in multiple ways.

The purpose of this paper is to illustrate the research design that supports the following key design and instructional features of the Zeos test preparation program:

- Alignment to Common Core State Standards.
- Progress monitoring for teachers.
- Proprietary mastery algorithms.
- Distributed practice.
- Incorporation of cognitive load theory.
- Incorporation of game-based learning design elements.
- Use of reinforcement and immediate feedback.
- Influence of peers on motivation.

Through the inclusion of these research-based features, the program aims to provide reliable and accurate information on student progress and to positively impact student achievement.
Zeos aligns to Common Core State Standards

*In a perfect world, what a student is tested on should be derived from what is expected of the student as detailed in the state and district standards, as well as from what is taught to the student by his or her teachers.* (Martone & Sireci, 2009, p. 1334)

With the advent of NCLB (2002), states across the country created their own standards for student learning. These standards, while useful, were varied, with studies suggesting that state-to-state alignment or agreement between standards hovered around 25% (Porter, Polikoff & Smithson, 2009). To establish a more coherent and unified set of standards, the National Governors Association Center for Best Practice and the Council of Chief State School Officers created Common Core State Standards. These core standards set expectations related to college and career readiness, and provided explicit learning objectives for students at each grade level in K–12, thus guiding teachers’ instructional goals. The standards are evidence-based and incorporate various state and international standards. As of May 2012, 45 states including the District of Columbia reported using Common Core State Standards in English language arts and math (National Governors Association and the Council of Chief State School Officers, 2012).

Once standards are developed, aligning high-stakes practice assessments and instruction to standards is considered to be best practice (Clarke, Stow, Ruebling, & Kayone, 2006; Decker & Bolt, 2008; Fraser & Kahle, 2007; Kim & Crasco, 2006). This allows for a comprehensive view of student knowledge of standards and strengthens the validity of assessments (Martineau, Paek, Keene & Hirsch, 2007; Martone & Sireci, 2009).

High-stakes assessment practice should be aligned to standards in such a way that assessments reflect both the range of content and level of knowledge required for mastery of skills (Decker & Bolt, 2008). When aligned in this way, teachers can use data from ongoing high-stakes practice assessments to identify strengths and weaknesses of student learning prior to larger state high-stakes assessments (Clarke et al. 2006), resulting in numerous opportunities to improve and modify instruction throughout the school year (Webb, 1997). Thus, the alignment of high-stakes assessment practice and standards enables students and teachers to learn and grow from a deeper understanding of student progress toward state expectations.

Teachers benefit when classroom instruction is guided by a map of how curriculum components align with standards (Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2008), as this deepens their understanding of how instruction supports expectations, frees them from time spent creating new instructional materials (Smith & O’Day, 1990), and allows them to see how curriculum and quality of instruction impact student achievement (Decker & Bolt, 2008).

**How Zeos Supports Standards Alignment**

Zeos provides standards practice using content that is aligned to state and Common Core State Standards in grades 3–10 English language arts and mathematics. By using a standards-aligned high-stakes assessment practice program, teachers can monitor student progress throughout the year and modify instruction to help students meet standards.
Other studies suggest that positive, standards-based classroom environments can mitigate potential negative effects of peer or family relationships on student achievement in math and science (Fraser & Kahle, 2007), that higher degrees of alignment between instruction and standards positively predicts reading achievement (Smithson & Collares, 2007), and that the use of a standards-aligned program results in greater math achievement gains when compared to classrooms not using a standards-aligned program (Hannafin & Foshay, 2008). By using a program that aligns high-stakes assessment practice with standards and informs instructional opportunities based on standards, teachers can hope to positively effect student achievement outcomes.

**Zeos provides progress monitoring for teachers**

*Consistent use of the teaching/learning cycle as a basis for instructional decision-making has a positive impact on student learning, promotes lasting change in teacher practice, and produces teachers and students who are empowered learners. (Jinkins, 2001, p. 283)*

Students benefit from classroom environments where teachers are continually seeking ways to improve student learning. One such environment incorporates the teaching/learning cycle in which teachers use student data to modify their instruction. In the teaching/learning cycle, teachers continually work to understand student skills and needs, and use this understanding to inform future lessons (Jinkins, 2001). When used effectively, the teaching/learning cycle becomes a continuous feedback loop that results in student progress.

One method for understanding current student performance is through the collection of progress monitoring data, which studies suggest can be beneficial for both at-risk (Fuchs & Fuchs, 2001) and not-at-risk student populations (Deno et al., 2009). Several studies have found positive effects of progress monitoring on student achievement (Baker, Gersten & Lee, 2002; Bolt, Ysseldyke, & Patterson, 2010; Jinkins, 2001; Stecker, Fuchs & Fuchs, 2005). For example, Jinkins (2001) conducted a study where teachers learned about the teaching/learning cycle and then used progress monitoring techniques in their classroom to make instructional decisions and to determine areas of student need. In these classrooms, students advanced 2–3 reading levels over six months to one year. Jinkins (2001) suggested this method enhanced student motivation and achievement because it helped teachers build upon what students already knew.

Similarly, a meta-analysis of the literature on the effects of math intervention on low-achieving students found that when teachers monitored student performance, graphed their progress, and shared that information with students, the students made significant achievement gains. When teachers actually modified their instruction based on student performance data, student achievement gains were even greater (Baker et al., 2002).

Research also suggests that computer programs can be used to monitor student progress, target areas of individual need and provide feedback to teachers and students (Baker et al., 2002). One study using weekly computerized progress monitoring found that students who received individualized instruction from teachers based on their performance achieved at
higher levels compared to students who did not receive individualized and performance-based instruction (Stecker & Fuchs, 2000). Thus, computers can provide frequent progress monitoring data to guide individualized instruction and support.

Perhaps one key to the effectiveness of progress monitoring is the feedback it provides to teachers. Ball and Gettinger (2009) found that when teachers do not receive feedback on student performance, students have lower levels of improvement, suggesting the importance of training teachers in the use of progress monitoring feedback. Taken together, research supports the value of incorporating progress monitoring strategies into the classroom through use of a teaching/learning cycle. When teachers follow these strategies, they gain a greater awareness of student progress toward standards and can modify their instruction accordingly to help students succeed.

Zeos assesses student mastery through use of proprietary mastery algorithms

The characteristics of behavioral decision-making processes that affect risk taking may shape people’s behavior during tests. Individual differences in examinees’ tendencies to take risks and the use of different scoring methods are likely to alter the patterns of responses systematically. This can provide advantages or disadvantages for certain groups of examinees, which may lower the tests’ validity. (Bereby-Meyer, Meyer, & Flascher, 2002, p. 313)

Traditionally, teachers grade multiple-choice tests by indicating the percent of correct items. However, by using percent correct methods, tests become influenced by chance and are increasingly unreliable (Burton & Miller, 1999), as students are likely to guess when they are uncertain of the correct response (San Martin, del Pino, & De Boeck, 2006). By its nature, the percent correct method cannot distinguish between correct answers from content mastery versus guessing, subsequently lowering test reliability (Bereby-Meyer, Meyer & Flascher, 2002; Burton, 2005). As a direct consequence, when teachers use percent correct as a grading method, they experience reduced accuracy in assessing student mastery of topics and are left with an incomplete picture of student understanding (Hoe, Kiong, Sam & Usop, 2009).
There are several factors that can influence guessing. First, risk takers may be more likely to guess on tests and to benefit from guessing, leaving risk-averse students at a disadvantage (Bereby-Meyer et al., 2002). Second, guessing is more likely to occur on later items in a test (Chang, Plake, Kramer & Lien, 2011). Third, guessing is more likely to occur on language arts than math tests, perhaps because of a greater ability to make educated guesses on language arts items (San Martín et al., 2006). Finally, students are more likely to guess when the test is longer (e.g., 50 vs. 25 questions) or they have lower final class scores (e.g., D vs. A) (Bereby-Meyer et al., 2002). As a result, there are numerous situations, student and testing characteristics that can increase the likelihood of guessing and subsequently reduce test validity and reliability.

Previous attempts to counterbalance guessing have used the deduction method, where points are both awarded for correct responses and deducted for incorrect responses. Burton (2001, 2005) argues that it is difficult to assess knowledge without using point deductions, as students who are on the borderline can end up passing an exam because of guessing. However, one study by Betts, Elder, Hartley, and Trueman (2009) suggests that there might be no benefit to the deduction method. These researchers found that students who were tested using percent correct scored higher on exams than students in the deduction method group, but when researchers later employed a correction for guessing in the percent correct group, there was no significant reduction in student scores. Interestingly, in the deduction method group, students answered fewer questions. As a result, percent correct, in contrast to the deduction method, provides students with opportunities to show partial mastery of content.

Another scoring technique attempts to correct for guessing by using program algorithms based on the statistical components of the questions themselves (Harper, 2003). For instance, algorithms could take into account the strategies employed by number right elimination testing, which asks students to eliminate incorrect items and gives points for correct eliminations. Students have found this type of grading to be helpful in learning math (Hoe et al., 2009) and one study found that guessing was reduced while employing this strategy (Lau, Lau, Hong, & Usop, 2011). The use of program algorithms can positively benefit teachers’ understanding of student mastery.
Zeos provides opportunities for distributed practice

More than 100 years of distributed practice research have demonstrated that learning is powerfully affected by the temporal distribution of study time. More specifically, spaced (vs. massed) learning of items consistently shows benefits, regardless of retention interval, and learning benefits increase with increased time lags between presentations. (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006, p. 371)

In learning new material, one fundamental question always remains: “How do we learn best?” Over several years of research into the nature of learning, researchers have examined the benefits of distributed (spaced) learning versus massed (cramming) learning and practice. Meta-analyses of the literature point to two key aspects of the power of distributed practice: first, that distributed learning results in greater retention of material (Donovan & Radosevich, 1999); and second, that distributed practice is effective in retention intervals as short as one minute and as long as 30 days (Cepeda et al., 2006). The benefits of distributed versus massed practice hold true in child, adolescent, and adult populations (Seabrook, Brown & Solity, 2005).

What makes distributed practice effective? Benjamin and Tullis (2010) suggest that distributed practice is beneficial in reminding the student of previous learning occasions. Therefore, initial learning is critically important because repeated sessions enhance memory for the original event (e.g., if you learn to ride a bike poorly the first time, then you will continue to practice those poor behaviors). However, these researchers acknowledge that they are unsure to what degree feedback on performance affects subsequent behavior. Roediger and Karpicke (2006) suggest that frequent and distributed testing with feedback can be useful in making future reviews of material more productive. Students not only practice on these occasions, but practice with corrected information. Roediger and Karpicke (2006) suggest this encourages students to study more, to be more engaged, and to experience lower levels of test anxiety. These theories are supported by one study finding that higher levels of review are associated with greater retention. Specifically, repeated practice (across 90 trials) in a distributed condition was associated with recalling twice as many items at a final one-week recall compared to massed practice (Karpicke & Bauernschmidt, 2011). When students distribute practice, they are given new opportunities to review material. With each practice and feedback session, learning is enhanced.

Much of the previous research examined how distributed practice helps students retain information over short periods (i.e., less than one month). However, in an educational setting with high-stakes assessments, students may be expected to retain information for several months to a year. What does research on distributed practice tell us about retention over longer time periods? In general, more time between a final practice session and testing is associated with poorer recall and recognition (Cepeda, Vul, Rohrer, Wixted & Pashler, 2008). To increase recall over the long term, Cepeda et al. (2008) suggest teachers use more spacing between review periods, called the “optimal gap.” The optimal gap for long-term retention of content over 7 to 360 days ranged from 1 to 21 days between practice sessions and was associated with a 64% increase in recall and a 26% increase in recognition. Cepeda et al. (2008) noted the optimal gap is not some absolute number, but suggest that if teachers want students to
remember something for a long time, students should distribute practice over an extended period of time. Similarly, one study emphasized the importance of distributing practice over several months, finding that students who spent six months on a topic versus eight weeks had higher levels of conceptual understanding and higher final course exam grades (Budé, Imbos, van de Wiel, & Berger, 2011). Spacing of information is critically important in long-term retention. If teachers want students to retain information over the long term, it is important to distribute practice and review over extended periods of time.

Mastery also plays a critical role in long-term retention of content. Research suggests that in order for information to become fully ingrained over the long term, it needs to be practiced repeatedly even after it is consistently performed correctly. In order to achieve this level, individuals need to have efficient distributed practice or review over an extended period of time (e.g., months rather than days or weeks) (Cepeda et al., 2009; Rohrer & Pashler, 2007; Schwartz, Son, Kornell, & Finn, 2011; Willingham, 2004). As an example, in one study examining the effects of mastery and overlearning on performance, Péladeau, Forget and Gagné (2003) assigned math students to one of four groups:

1. Mastery only (students practiced a unit until they achieved 85%, after which point mastered quizzes were removed for the rest of the semester);
2. Accuracy-oriented overlearning (same mastery criterion of 85%, but students were told to keep practicing beyond this point two times per week over 5 weeks and to focus on increasing their accuracy);
3. Fluency-oriented overlearning (same as accuracy-oriented group, but students were told to focus on increasing the number of correct responses with each practice); and
4. No mastery control group (used the same practice software but did not practice until the point of mastery).

Five to six months after students stopped reviewing math content, researchers contacted participants for a long-term retention test measuring student recall and transfer. Péladeau, Forget and Gagné (2003) found that students in the overlearning groups scored significantly higher than those in the mastery-only condition or in the no mastery control group, and students in the mastery only group scored higher than students in the non-mastery control group. The study demonstrates the importance of not only practicing material, but also practicing it to the point of mastery or overlearning in order to have long-term impacts on retention.
Zeos incorporates elements of cognitive load theory

**Good instructional design is driven by our knowledge of human cognitive structures and the nature in which those structures are organized into a cognitive architecture. Without knowledge of relevant aspects of human cognitive architecture such as the characteristics of and intricate relations between working memory and long-term memory, the effectiveness of instructional design is likely to be random. (Sweller, 2005, p. 19)**

Every time we learn a new topic, that information is first reviewed by our working memory and (hopefully) transferred to long-term memory. Unfortunately, we know that working memory is a transitory holding center with limited processing capacity (Mayer, 2005; Mayer & Moreno, 2003; Sweller, 2005). The mind is not simply a video recorder that can replay an exact image of what it sees or hears. Rather, individuals select and process information in order to make sense of it (Mayer, 2005). The level of mental activity required to process information in working memory is known as cognitive load (Haslam & Hamilton, 2010; Sweller, 2005). In general, the more complex the task, the greater the cognitive load, and the more difficult the task is to learn.

To optimize cognitive load while learning, we must transfer new information from working memory to long-term memory. Information that is rehearsed is more likely to successfully make this transition, while other material is discarded (Mayer, 2005). The use of schemas also helps. Sweller (2005) suggests that schemas, or information organizers (e.g. all birds have wings), allow multiple types of information to be grouped together. After acquiring a schema, new information that fits within that schema can be processed automatically, without conscious awareness. This automatic process allows the individual to focus on other areas of importance. Schemas and automaticity help direct how information will be processed and retained in long-term memory and thus reduce the cognitive load on working memory.

Instructional design recognizes three types of cognitive load—extraneous, intrinsic, and germane (DeLeeuw & Mayer, 2008). Extraneous cognitive load taxes the limits of working memory and does not involve the creation of schemas (Sweller, 2005). Mayer and Moreno (2003) also referred to this as incidental processing, or information not required for understanding but that arises because of the nature of the instructional presentation (e.g., poor designs or distracting layouts). Because this type of cognitive load is outside the nature of the task, Sweller (2005) suggests finding ways to minimize this type of load.

Intrinsic cognitive load is inherent to the task at hand (Sweller, 2005). Mayer and Moreno (2003) referred to intrinsic load as essential processing, or thought that is required to make sense of the material. For example, intrinsic load in foreign language learning would deal with learning new vocabulary and understanding how grammar rules apply in this new situation.

Germane cognitive load is caused by learning and results in the creation of schemas. Instruction that includes multiple examples of a single idea can increase germane cognitive load and supports the creation of schemas. The total cognitive load required by a task is the sum of all the types that are involved (Sweller, 2005). Cognitive overload occurs when total processing
exceeds learner capacity, at which point the learning process stalls (Mayer & Moreno, 2003; Sweller, 2005).

Based on an understanding of the different types of cognitive load, researchers have suggested several ways educators can reduce the risk of cognitive overload:

- Offer schemas for organizing information, (Sweller, 2005).
- Present information in small pieces rather than all at once (Mayer & Moreno, 2003).
- Link new lessons to previously learned material (Mayer & Moreno, 2003).
- Eliminate redundant information (e.g., if there are captions in an animation, they should not also be spoken) (Mayer & Moreno, 2003).
- Direct the learner to the important parts of the lesson (Mayer & Moreno, 2003).
- Synchronize related content so that it is presented together (e.g., text and illustrations on one page) (Haslam & Hamilton, 2010; Mayer & Moreno, 2003).

When educators understand how working memory processing relates to long-term retention and adjust their instructional practices based on these concepts, student learning is enhanced.

**HOW ZEOS INCORPORATES ELEMENTS OF COGNITIVE LOAD THEORY**

Zeos has several platform and design features aimed at reducing extraneous cognitive load, optimizing intrinsic cognitive load and maximizing germane cognitive load.

The user interface and challenges focus on the task at hand and do not have extraneous details embedded in the design. When students complete challenges, they only are concerned with the question on the screen, possible answer choices, and the learning objective associated with the challenge. Similarly, the entire platform is clean, easy to understand, and quickly navigable.

Zeos supports intrinsic and germane cognitive load through quizzes or challenges. Questions are clear and easy to understand and directly apply to state and Common Core State Standards. When students experience difficulty with a question, Zeos provides additional feedback, thus helping students to build schemas for concepts.
Zeos incorporates elements of game-based learning design

The generally accepted position is that games themselves are not sufficient for learning but that there are elements of games that can be activated within an instructional context that may enhance the learning process. (Garris, Ahlers, & Driskell, 2002, p. 443)

How do we motivate students to learn something new? Researchers suggest that teachers can encourage intrinsic motivation (i.e., valuing an activity for its own sake) by offering students the experiences of choice, control, and feelings of competency in a skill. Students are motivated by feeling like they can achieve a goal and have a positive outcome (Deci, Eghrari, Patrick & Leone, 1994; Ryan & Deci, 2000). Similarly, research on the concept of flow (i.e., being immersed in an intrinsically motivating activity) suggests that when students experience high levels of challenge and competence, they report higher levels of engagement. Additionally, perceptions of control are associated with higher levels of engagement, esteem, and mood (Shernoff, Csikzentmihalyi, Schneider, & Shernoff, 2003). These core components might explain why individuals find games to be motivating, as they frequently offer opportunities for choice, challenge, and feelings of competency.

Across several studies, researchers have identified some common characteristics of game-based learning that encourage student engagement, including: challenges, fantasy, curiosity and uncertainty, scaffolding, interaction, control, and opportunities to try out different characters.

Challenge. Malone (1981) notes that games need to have some sort of challenge or an attempt to achieve a goal. Students prefer the right amount of challenge and clear goals, which can lead to greater attention, motivation, and engagement (Garris, Ahlers & Driskell, 2002). In one study, researchers found children derived more enjoyment out of mastering difficult compared to easy anagrams (Harter, 1974), suggesting that challenge is key to fostering engagement and interest. Programs that create challenging environments with clear expectations can be beneficial.

Fantasy. Fantasies increase retention of content because they can generate vividly imagined scenes that become entrenched in memory (Malone, 1981). Imagined worlds and fantasies embedded within learning are also more likely to make content interesting and motivating (Garris, Ahlers & Driskell, 2002).

Curiosity and Uncertainty. Games should have some expectations but also some uncertainties. Students prefer some level of mystery (Garris, Ahlers & Driskell, 2002; Malone, 1981). Curiosity can be enabled by challenging students’ existing schemas as incomplete or inconsistent, motivating them to reconcile inconsistencies (Malone, 1981). Students also prefer situations with uncertain reward outcomes. In one study by Howard-Jones and Demetriou (2009), students were more likely to choose an outcome with a 50% chance of no reward and a 50% chance of a higher reward compared to a 100% chance of a lesser reward. Encouraging curiosity and uncertainty are two ways to pique student interest.
Scaffolding. Game-based learning is most effective when students have the support of another person (teacher, parent, or peer) and material is reinforced by other instructional methods (An & Bonk, 2009; Young et al., 2012). Feedback on progress toward achieving goals can also help to enhance performance and increase efforts toward mastery (Garris, Ahlers & Driskell, 2002). This support structure, or scaffolding, is a key element of effective instructional games.

Interaction. Allowing for interaction between players and between learners and the teacher is important (An & Bonk, 2009). Research shows that adolescents and adults prefer multiplayer games because of the social interaction and opportunity to compete with one another (Trespalacios, Chamberlain, & Gallagher, 2011; Yee, 2006). Additionally, students can benefit from collaborative and competitive learning environments, resulting in greater learning when they embrace the competitive aspect (Admiraal, Huizenga, Akkerman, & ten Dam, 2011). Peer and teacher interaction with students serves as one vehicle for increasing interest and learning outcomes.

Control. Similar to the research on motivation, feelings of control over learning or a sense of ownership are important in games (An & Bonk, 2009). Feelings of control lead to more positive feelings toward games, greater motivation and learning (Garris, Ahlers & Driskell, 2002).

Trying out different characters. Games offer the ability to take on a different persona or role (An & Bonk, 2009) and research suggests use of avatar characters gives students feelings of ownership over their own learning (Falloon, 2010).

When teachers incorporate elements of games into their instruction, classrooms can benefit from increased levels of engagement and motivation.

HOW ZEOS INCORPORATES ELEMENTS OF GAME-BASED LEARNING DESIGN IN THE PROGRAM

When students first sign into Zeos, they create a superhero avatar that stays with them throughout the program. By successfully completing challenges, students have the opportunity to get more items for their superhero. Students are also introduced to test-taking strategies through superhero academy comics that immerse the student in a fantasy world.

Each challenge in Zeos has both a superhero goal and a knowledge goal. Students are encouraged to correctly answer at least 50% of questions to help a superhero with a task, and if they correctly answer 100% of questions, they have the opportunity to spin for a reward. Rewards are uncertain and can vary with the spin of a wheel. Knowledge goals are always posted at the top of each challenge and relate to state and Common Core State Standards.

Students have a choice in the challenges they decide to complete and can complete a new challenge or one they have taken previously. Students also receive scaffolding through use of immediate feedback on incorrect answers.

In addition, the game allows for peer interactions, creating an environment where students can feel connected to one another and to the Zeos world. As a result, the program aims to help students feel more invested in the experience and more motivated to spend time in Zeos.
Zeos provides students with reinforcement and immediate feedback

*Research in classroom instruction has demonstrated that feedback can be a powerful instructional variable for improving student learning.* (Robinson, DePascale, & Roberts, 1989, p. 28)

Reinforcement and feedback are important learning tools in the classroom. Numerous research studies have shown how the neurotransmitter dopamine is involved in reward and reinforcement in the brain (e.g., Cannon & Bseikri, 2004; Fiorillo, Tobler & Schulz, 2003). At a 50% probability of reward, dopamine spikes at higher levels compared to when there is a 100% chance of reward. This increase is also associated with the size of the reward, with larger rewards linked to a greater release of dopamine (Fiorillo, Tobler & Schulz, 2003).

The spike in dopamine in response to uncertain situations with high payoff may explain different studies of educational risk taking. For instance, consider the study by Howard-Jones and Demetriou (2009) described in the gaming section, where students were more likely to choose an uncertain outcome with a greater reward than a certain outcome with a lesser reward. One study also suggests students might attempt more difficult problems in order to gain a larger reward as opposed to trying easier problems that are associated with a lesser reward (Clifford & Chou, 1991). As a result, variable or uncertain reinforcement schedules might lead to increased engagement or interest in activities.

Feedback is also a critically important tool in learning. A wealth of research emphasizes a number of ways in which feedback promotes learning. First, researchers find that achievement and retention is higher in conditions where students receive any type of feedback compared to no feedback (Azevedo, 1995; Pashler, Cepeda, Wixted, & Rohrer, 2005). Second, immediate feedback is more effective at improving student performance than delayed feedback (Butler & Roediger, 2008; Epstein et al., 2002; Kulik & Kulik, 1988). Third, the type of feedback matters. Task feedback involves basic information about the task (e.g., item is correct or incorrect) and is helpful when students do not have the knowledge required by the task. Process feedback is aimed at the process necessary to complete a task (e.g., suggesting that students use more descriptors in an essay). This type of feedback is helpful for encouraging a deeper level of learning compared to task feedback. Self-regulation feedback involves encouraging students to be self-reflective in their own learning and confident in a specific task. In contrast, self feedback is unrelated to performance on a specific task and is synonymous with praise (e.g., “You are a great student!”). This type of feedback is used too often but is ineffective because it is not directed to the task at hand (Hattie & Timperely, 2007).

Feedback that provides students with the correct answers and avoids praise are related to stronger effects on performance (Kluger & DeNisi, 1996). In addition, students who receive frequent self-regulation feedback from peers and adults have enhanced levels of self-concept and attribute their success more to effort compared to students who receive less frequent and self-regulated feedback (Craven, Marsh & Debus, 1991).
Hattie and Timperley (2007) suggest effective feedback methods include providing goals for students, conveying success toward meeting goals, and providing information about student progress. Effective feedback can reduce misinformation and increase retention. For example, if students receive feedback on an item, they are less likely to choose an incorrect answer a second time (Butler & Roediger, 2008; Epstein et al., 2002). In sum, students benefit from feedback that is supportive of student learning, progress, and retention; that acknowledges the requirements of the task; and is not too focused on praise.

Computers have been shown to be useful tools for effective feedback (Azevedo, 1995; Baker et al.; Robinson et al., 1989), especially when that feedback focuses attention on the task (Kluger & DeNisi, 1996). Thus computer interventions that provide task and self-regulated feedback can have positive effects on achievement.

**HOW ZEOS SUPPORTS REINFORCEMENT AND FEEDBACK**

In the Zeos program, students are rewarded for mastery of a concept. When students master a challenge, they are given the opportunity to spin a reward spinner that provides different access to premium items for their superhero avatars. The size and type of the reward operate on a schedule of variable reinforcement, with the type of reward being left up to chance.

Zeos provides students with immediate feedback on quiz performance, offering correct information on missed questions and indicating areas where improvement is needed. Students who continue to have difficulty receive more factual feedback in the problem area, but are not cycled down to a lower level.

By allowing students to continue working at the assigned grade level, the program offers potential benefits. First, students have continued opportunities to practice with grade-level assessment items, which help in preparing for high-stakes assessments. Second, students have the opportunity to learn from their errors through corrective feedback and to improve their understanding of key concepts.
Zeos acknowledges the influence of peers on motivation

*Feeling connected and important is not just a by-product of doing well in school; a sense of belonging or relatedness plays an integral role in children’s motivational development. (Furrer & Skinner, 2003, p. 161)*

The supportive role that peers play in our lives can be a protective factor in times of stress (Dubow & Tisak, 1989). In the classroom, peer support can generate feelings of belonging and connectedness helping students persist longer at tasks and experience higher levels of motivation (Walton, Cohen, Cwir & Spencer, 2011). In addition, feeling connected to peers and teachers in the classroom is a significant positive predictor of emotional engagement in class content (Furrer & Skinner, 2003).

Peers also serve as a source of support and feedback when discussing classroom experiences. One study by Styers, Baker-Ward, and Turner (2009) found that elementary and middle school students talk more with their friends about satisfying compared to disappointing academic experiences. Students might be more likely to share positive experiences because these bolster positive attitudes, as evidenced in a study by Altermatt (2011). In this study, children reported more positive attitudes toward school when they talked about their successes with peers—either because they were happy about the experience of communicating their accomplishments, or to demonstrate that school is important to them. Elementary and middle school students also discussed academic successes with peers for different reasons, with younger children sharing because they were happy or because they saw school as important, and older children sharing to compare themselves to others. Successful experiences in academics are something that students of different ages enjoy sharing with their peers.

Friends can also support prosocial relations with others. Friendship quality and social support both serve as positive predictors of school engagement, whereas aggression is a negative predictor of school engagement. These findings suggest the importance of maintaining positive peer relationships and minimizing aggressive tendencies to promote school engagement (Perdue, Manzeske, & Estell, 2009). Prosocial video games can help promote positive social behavior. Studies find that students who play prosocial video games are more likely to help an unknown peer or to help an experimenter compared to students who play aggressive or neutral video games (Gentile et al., 2009; Greitmeyer & Osswald, 2010). In an increasingly technological world, peers and friends have multiple opportunities and platforms on which to interact. Seventy-three percent of teens now use social networking sites and 66% of teens send and receive text messages (Lenhart, Purcell, Smith & Zickuhr, 2010). Students are communicating more now by text than in person or by phone (Timmis, 2012), and are increasingly finding ways to share ideas with one another online (Kitsis, 2008). Consequently,
peers can be a supportive and motivating force without being there in person.

Conclusions

New benchmarks for academic accountability, such as Common Core State Standards, aim to support K–12 students’ ongoing learning as well as college and career readiness (National Governors Association and the Council of Chief State School Officers, 2011). Programs that align with these standards can not only influence student achievement in the short term, but also have lasting impacts throughout students’ lives. The notion of a common set of standards also is supported at a national level through the use of No Child Left Behind waivers given to states committed to improving standards and student progress monitoring (Duncan, 2012).

Zeos is a progress monitoring and state high-stakes assessment practice program for students in Grades 3–10. Its program design is supported by key findings from research in number of areas:

Alignment to Common Core State Standards. Alignment practices are critical for understanding student progress toward standards achievement.

Progress monitoring for teachers. Progress monitoring is an effective strategy for promoting teacher awareness of student performance and suggesting areas where students need additional support.

Algorithms to more accurately determine mastery. The incorporation of proprietary mastery algorithms gives teachers a more accurate and reliable sense of student mastery of content.

Distributed practice. Distributed practice gives students new chances to reinforce information and provides opportunities for mastery and overlearning, both of which support long-term retention.

Cognitive load theory. Cognitive load theory suggests that certain types of information make learning difficult, while other strategies (e.g., schemas) are supportive of learning and long-term retention.

Elements of game-based learning design. Researchers have described a variety of ways in which video games motivate and engage students.

Reinforcement and immediate feedback. Variable reinforcement and immediate feedback are important tools for learning.

Peer influence of peers on motivation. Peers have a unique influence on students, serving as a positive source of support in a wide variety of situations and environments.

Zeos provides a unique approach to test preparation grounded in a strong research design.
References


Notes

1 Minnesota is not officially recognized on the current list because they only accepted the English language arts Common Core State Standards. In addition, Hawaii, Nebraska, Texas, and Virginia have not yet adopted the Common Core State Standards.

2 Fraser and Kahle (2007) conducted a study with over 3,000 students in 392 classes, examining the impact of classroom and peer factors on student achievement. Calculations from regression coefficients translated to effect sizes of 0.87 to 1.50 for the effect of standards-based classroom practices on student achievement.

3 Smithson and Collares (2007) conducted a regression model and found that when controlling for grade level and prior achievement, degree of alignment positively predicted reading number correct (effect size = 31%) and reading scale score (effect size = 57%).

4 Hannafin and Foshay (2008) compared students who participated in standards-aligned, computer-based instruction in math to students who did not receive standards-aligned, computer-based instruction in the same district. The study found that the computer-based instruction group had significantly greater improvement in math scores compared to the other group (effect size = 1.27).

5 Baker, Gersten, and Lee (2002) conducted a meta-analysis of 15 studies on interventions designed to help at-risk students in math. The researchers found an effect size of 0.29 for the impact of progress monitoring and sharing progress monitoring data on student achievement, and an effect size of 0.51 for using information on student performance to adapt instruction.

6 Students in grades 2–8 completed curriculum-based measurement (CBM) computer assessments two times per week for 20 weeks. The software calculated scores and showed a graph of performance. When teachers modified instruction based on individualized student performance compared to a group whose instruction was modified based on the needs of other students in the classroom, students with individualized and performance-based instruction saw greater achievement on a math test (effect size = 0.87) (Stecker & Fuchs, 2000).

7 Ball and Gettinger (2002) found that students in the feedback condition compared to the no feedback condition scored significantly higher on letter naming fluency ($\eta^2 = .13$), initial sound fluency ($\eta^2 = .06$) and phoneme segmentation fluency ($\eta^2 = .08$).

8 Bereby-Meyer, Meyer, and Flascher (2002) conducted a study with college students on confidence in guessing behavior and found that college students are more likely to guess when exams are longer versus shorter (effect size = 0.34) and when they expect lower versus higher final class grades (effect size = .18).

9 Betts, Elder, Hartley, and Trueman (2009) found that students without a correction for guessing scored higher on exams (partial $\eta^2$ ranged from .06 to .24), and those with a correction for guessing answered significantly fewer questions (partial $\eta^2$ ranged from .25 to .38). When researchers later employed point deductions to the percent correct group student scores, there was no significant reduction in scores.

10 Hoe, Kiong, Sam, and Usop (2009) created a percent correct elimination testing software for math assessment for Malaysian students. The study found that 64% of students felt the software was helpful for learning math.
In a meta-analysis of the literature, Donovan and Radosevich (1999) found a weighted mean effect size of 0.46 in favor of distributed learning.

Cepeda et al. (2006) conducted a meta-analysis of the distributed practice literature using 184 articles. The researchers found a large positive benefit of spaced versus massed practice for retention intervals of 8–30 days (calculated effect size = 1.45) and for retention intervals of one minute (where students in a distributed practice condition scored 9% higher).

Seabrook, Brown and Solity (2005) conducted several experiments with students six years old to adult and found strong effect sizes (ranging from 0.34 to 1.42) in favor of distributed over massed practice.

Karpicke and Bauernschmidt (2011) asked 96 undergraduates to learn Swahili-English word pairs and had four groups of learning conditions—short spacing (practice over 15 trials), medium spacing (practice over 30 trials) and long spacing (practice over 90 trials)—and tested them one week after learning trials ended. Medium spacing produced greater long-term retention than short spacing (64% vs. 49%) and long spacing was more effective than medium and short spacing (75% vs. 64% and 75% vs. 49%).

Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) taught 1,354 participants facts to the point of mastery. Then students had a gap between mastery and review sessions of 0 to 105 days. After the second review session, students had a retention interval period of 7 to 350 days, after which point students completed a recall and recognition test. The researchers found the optimal gaps (time between review sessions) for retention intervals (time between final review and final test) were as follows for recall performance: 7 day retention interval, 1 day gap; 35 day retention interval, 11 day gap; 70 day retention interval, 21 day gap; 350 day retention interval, 21 day gap. The optimal gap was associated with a significant increase in recall ($d = 1.10$) and a significant increase in recognition ($d = 1.5$).

Budé, Imbos, van de Wiel, and Berger (2011) followed groups of students before and after a change to length of curriculum. The researchers found that students in the 6-month vs. 8-week course had significantly higher levels of conceptual understanding on the topic (effect size = 1.10) and significantly higher final course exam grades (effect size = 0.61).

Cepeda, Coburn, Rohrer, Wixted, Mozer, and Pashler (2009) taught 215 college students Swahili-English word pairs to the point of mastery. The researchers had two experiments where students had a 0 to 7 day gap between practice sessions and a testing 10 days after the final practice session, or 0 to 168 days between practice sessions and testing 6 months after the final practice session. Cepeda et al. (2009) found that a 0 day gap between learning sessions resulted in significantly worse performance than gaps of 1 to 7 days. Additionally, for the 6-month test delay, a 28-day gap was considered optimal.

Péladeau, Forget, and Gagné (2003) found that those in the overlearning groups scored 0.42 to 0.57 standard deviations greater than students in the mastery only condition. Total retention of the non-mastery group was 55.7% compared to 65.3% in the mastery group ($d = 0.77$). Students in the overlearning groups also performed better than the mastery groups with retention scores of 75.2% and 75.0% (overlearning group means were not significantly different).

Deci, Eghrari, Patrick, and Leone (1994) gave students the non-interesting task of pushing a spacebar when seeing lights, and found that engagement time increased linearly by whether there was a rationale for the task, acknowledgement of the participant’s view, and whether participants believed they had a choice in the activity (calculated effect size = .31). In addition, participants believed experiments in which they were provided a rationale for the activity were more enjoyable (calculated effect size = .51).
Shernoff, Csikzentmihalyi, Schneider, and Shernoff (2003) gave high school students beepers and asked students to keep a log of their experiences each time the beeper went off over a one-week period. Students reported being engaged in instruction 73% of the time when they were in a state of flow (high challenge, high skill) versus a state of apathy (low challenge, low skill), during which their instructional engagement was 42%. Students also reported higher levels of engagement, esteem, and mood when given high versus low control (calculated effect sizes ranged from 0.47 to 0.97).

Harter (1974) asked children to solve anagrams, and looked at smiling in addition to perceptions of enjoyment. Children smiled more after solving more difficult anagrams (calculated effect size = .44) compared to easier ones, and 53% of students said they enjoyed figuring out difficult anagrams the most.

Elementary students played a game in which they were asked to answer whether a math fact was true or false. If they chose a certain outcome, they got 1 point for a correct answer. If they chose the uncertain outcome, they had an equal chance of getting 0 points or 2 points for a correct answer. Students preferred to choose the uncertain outcome to the certain outcome (p = .04) (Howard-Jones & Demetriou, 2009).

Clifford and Chou (1991) conducted a study with fourth graders and found that when participants were in a variable payoff condition, students chose more difficult problems and showed greater risk taking (calculated effect size = 2.46).

Azevedo (1995) found an overall weighted effect size of 0.80 across 22 studies in favor of the impact of computer-provided feedback versus no computer-provided feedback on achievement.

Pashler et al. (2005) found that when students received feedback on the correct answers on an assessment (versus no feedback), retention of material increased by 494% over a one-week period.

Butler and Roediger (2008) found that immediate feedback in a computerized testing condition led to more correct responses than no feedback (d = 0.69). Additionally, those without immediate feedback made more errors compared to those who received immediate feedback (d = 0.65)

Epstein et al. (2002) gave immediate feedback versus no feedback. After eight days, participants who received feedback answered more questions correctly compared to the no feedback group (p < .05). Those with no feedback answered significantly more questions incorrectly (p < .05) over the same time period.

Kulik and Kulik (1988) in their meta-analysis of the literature (n = 53 studies) found that immediate feedback is more effective than delayed feedback (effect size = .28).

Kluger and DeNisi (1996) in their meta-analysis of the feedback intervention literature found that providing the correct answer (compared to providing no information) is associated with stronger effects (effect size = 0.43). In addition, incorporating praise in feedback has a weak effect on performance (effect size = 0.09), but those types of feedback that avoid praise have a stronger effect (effect size = 0.34).

Craven, Marsh, and Debus (1991) compared students who received frequent feedback on their strengths and abilities to students who received less frequent and self-regulated feedback from teachers.
Those in the first condition saw an enhanced self-concept (effect size = 0.43) and attributed their success more to effort (effect size = 0.40) compared to students in the second condition.

31 Baker, Gersten, and Lee (2002) conducted a meta-analysis of 15 studies and found an effect size of 0.57 in favor of feedback recommendations from teachers or computers compared to no feedback conditions.

32 Kluger and DeNisi (1996) conducted a meta-analysis of the feedback intervention literature across 131 articles. Computer feedback was related to stronger effects on performance than non-computer feedback (effect size = 0.41).

33 Robinson, DePascale, and Roberts (1989) found an effect size of 0.40 in favor of computerized feedback compared to no feedback.

34 In four experiments, Walton et al. (2011) manipulated feelings of social connectedness to a stranger. When students felt more connected to the stranger, they persisted longer at tasks (effect sizes = 0.78–0.80) and expressed greater motivation for the subject area (effect sizes = 0.47–0.92).

35 Furrer and Skinner (2003) conducted a study with children in grades 3–6. They found that children’s perceptions of relatedness to peers and teachers significantly predicted classroom emotional engagement (explaining 12% and 5% of variance, respectively).

36 Styers, Baker-Ward, and Turner (2009) interviewed 270 participants (elementary, middle school, and college) about their memories for positive and negative academic experiences. All participants talked more with their friends about satisfying compared to disappointing academic experiences ($\eta^2 = .07$), with the elementary and middle school age groups talking more than college students about satisfying versus disappointing experiences.

37 Altermatt (2011) looked at peer relationships following academic success and interviewed 293 students in grades 5–8. When children shared success with peers either because they were happy or to show interest in school, students had more positive attitudes toward school (calculated effect sizes = 0.28 to 0.30). Younger children were more likely to share success with peers because they were happy (calculated effect size = 0.37) or perceived school as important (calculated effect size = 0.31) compared to older children. In contrast, older children were more likely to share success with peers because they were trying to compare themselves to others (calculated effect size = .26) compared to younger children.

38 In a study with college undergraduates, students played a prosocial video game (e.g., Super Mario Sunshine) or an aggressive or neutral game. Those in the prosocial game condition were more helpful to an unknown peer following the exercise ($d = 0.48$) (Gentile et al., 2009).

39 Greitemeyer and Osswald (2010) found that college students who played a prosocial video game were more likely to pick up pencils dropped by an experimenter, were more likely to intervene in an aggressive argument between a boyfriend and girlfriend, and were more likely to give more time to help with future experiments ($d = 0.82$) compared to those who played aggressive or neutral video games.