Game Design for Computer Science Education

Simon McCallum
Gjøvik University College

Abstract

Computer science students come from the gamer generation. A generation accustomed to receiving instant feedback and constant reward for action. This expectation creates tension between traditional assessment models and the learning style of the gamers taking our courses.

In this paper we discuss how game design principles can be used as part of modern instructional design. We then present two case studies at different levels of course design, from the design of exam assessment to the learning outcomes for a whole degree.

The first case study comes from a first year game design course where a confidence based multi-choice system was designed using game balance mechanics. We show that by using a game design approach we can create well balanced confidence intervals.

The second case study focuses on the use of experience points (XP) and character sheets to track student development and present learning objectives in a form familiar to our students. This can also be used to motivate extra curricular activity and organise group activities.

Objective

The goal of this paper is to investigate the use of game design principles in the development of assessment for tertiary courses, with specific reference to computer science courses. We assess the use of game design principles for instructional design, and propose a process for including game mechanics and design principles in the design of a whole degree program.

Background

Instructional design and game design share many traits. Instructional designers are trying to create an engaging educational experience where the end result is that the learner has developed new skills or knowledge. Game designers are also creating experiences, primarily with the objective of entertaining the player. The techniques used to entertain and engage the player often require an ever increasing level of difficulty. This engagement is often referred to as "flow" [Csìkszentmihàlyi, 1990]. For a player to experience flow the difficulty of the game must match the skill of the player. As the player improves at the game, it must respond by providing

This paper was presented at the NIK-2010 conference; see http://www.nik.no/.

more complex challenges. The process of mastering increasingly complex problems is analogous to the goal of instructional design, where the end result, rather than the process is usually emphasised.

Computer science education is very hierarchical in nature. Each course builds on experience gained in previous courses. Even though the specific order of topics covered is not agreed upon [Robins, 2010], the principle that each section is built on knowledge gained from the previous section is consistent. This is very similar to the process of learning to play many complex computer games, where starting in the middle of the game would be impossible as the skills needed have not been learned.

One of the main themes of computer science conferences is "how to increase the 'through put' of computer science courses" and "how to keep students motivated and engaged in the curriculum". Experiments with pair programming and adding game content Corney et al. [2010] tend to be successful in improving student retention. Generally these systems increase the quantity and speed of feedback to the students.

Game design is focused on what the player does and how they get feedback for actions. Both Gee [2004] and Prensky [2007] build convincing arguments that games are extremely effective learning environments. Gee claims that the flow state and player learning are part of all "good" games, and further, that the very structure of good games are well aligned with modern learning theory. Both systems advocate fast feedback, cycles of challenge followed by mastery, pacing to match player development, and well designed reward structures.

Context

Our students now spend more time playing computer games than reading books [Prensky, 2001, Statistics Norway]. Their appetite for instant feedback and constant rewards has increased as computers have been able to not only respond to the users' actions, but also connect players in massive social networks. Rather than making appointments and waiting for other people to become available, we can play games any time of the day or night with hundreds or thousands of other players. This, combined with mp3 players, digital video recorders, and the Internet, creates an environment where the individual determines the pace of interactions and when they occur. Although changes are occurring in the instructional environment at universities, the pace of change cannot keep up and so our educational environments are becoming increasingly out of sync with the natural tendencies of our audience, the gaming generation [Prensky, 2007].

Both instructional design and game design have long histories. Games have been used as part of learning environments for hundreds of years. Recently however, there has been a divergence of these two activities. One of the significant differences between game design and instructional design is the number and diversity of games that are being created. Thousands of digital games are published each year and most of them will not be profitable [Irwin, 2008, AndroLib, 2010, FADE, 2010]. The game development environment is much more competitive and less forgiving than academic course development. Failure is a constant threat to games companies, and the process of game design is critical to the companies survival. In the current academic environment the term "publish or perish" is used to refer to the measuring of research output. Instructional quality is much harder to measure and, even when courses are seen as weak, the consequences to the instructor are often negligible.

The competitive nature of the game development industry has forced game designers to be both innovative and reflective. Game mechanics that do not work well are quickly detected and discarded. Games that make it to the open market also provide test cases for new mechanics. If the core mechanic is poor the game is likely to fail. This strong feedback mechanism is more effective at improving game mechanics than the weak feedback provided for instructional design.

Case Study: Game Design Introduction

The first case study we present is focused on the game design principle of **balance**. We use this in two ways, firstly, as a tool for designing a multi-choice (MC) exam, and secondly, as part of the content of the game design course for which the exam was developed.

Multi-choice exams are widely used in tertiary education, and there is often debate about the marking system to use. Frandsen and Schwartzbach [2006] presents an analysis of marking systems and proposes a solution which balances knowledge vs guessing:

$$S_{axioms}(k,a,c) = \begin{cases} 0, & \textit{if } a = 0 \vee a = k \\ \log \frac{k}{a}, & \textit{if } a > 0 \wedge c = 1 \\ -\frac{a}{k-a} \log \frac{k}{a}, & \textit{if } a > 0 \wedge c = 0 \end{cases}$$

where k is the number of possible options, a is the number of checked options, and c is a Boolean indicating whether one of the checked options is the correct one. In the special case of there being one correct answer per question and each question having the same number of distracters, this results in the standard $+1$ for a correct answer and $-\frac{1}{k-1}$ for incorrect answers.

To create a strategic component to the answering of MC questions, we added a confidence factor to each question. The student is expected to answer the question and to indicate their level of confidence. Each confidence level is allocated a different marking system. The highest level of confidence should give the most marks, but should also carry a large cost to prevent choosing high confidence from being the dominant strategy.

Kolbitsch et al. [2008] present an analysis of a confidence based marking system. Their study uses four levels of confidence: Category 1 = 100%, Cat. 2 = 66%, Cat. 3 = 33% and Cat. 4 = 0%. This follows the general principle that students who are correct and confident receive the highest marks. There are two main problems with the study: one, the numbers were not disclosed to the students, so they could not know the relationship between confidence and the marking system; and two, the students responses were skewed strongly toward full confidence, thus meaning little could be gained from analyzing the small number of other answers.

Game Balance

To have meaningful interactions with a game, the player must be able to influence the game world. These interactions could be physical skill based, such as first person shooters, or cognitive decision making found in strategy games. For games to be interesting beyond the first interaction, the strategic choices must be balanced such that there are no dominant strategies.

The optimal strategy for selecting a marking category can be calculated as

$$R = p * C_1 + (1 - p) * C_0$$

Cat.	C_1	C_0	25%	50%	75%	100%
1	10	-5	-1.25	2.5	6.25	10
2	7	-3	-0.50	2.0	4.50	7
3	3	-2	-0.75	0.5	1.75	3
4	0	0	0	0	0	0

Table 1: Average expected return R for each category given a percentage likelihood of being correct using Kolbitsch et al. [2008] marking values.

Cat.	C_1	C_0	25%	50%	75%	100%
1	2	-2	-1.00	0.0	1.0	2.0
2	1.5	-0.5	0.00	0.5	1.0	1.5
3	1	0	0.25	0.5	0.75	1.0

Table 2: Marking system for IMT1361, with balanced values in each category, showing the average expected return R.

where p is the probability of getting the answer correct, C_1 the value of a correct answer and C_0 is the value of an incorrect answer. Table 1 shows the expected return R, for different marking categories vs. the probability of being correct. If the player is more that 40% certain about the answer they should select category 1 (100% confident 10,-5). Between 30% and 40% category 2 is the best option and below 30% category 4(guessing 0,0) is the best option. At no stage is category 3 (3 , -2) the best choice.

The correct labels for the categories given these marking schemes should be 40-100%, 30-40%, 0-30%. Given that there were only 5 possible answers, the lowest value could be replaced with 20%. Therefore most of the time (75%) students should select the original 100% category. And although they did not have access to these numbers, more than 90% of the time the students selected the total confidence category.

Multiple choice marking for IMT1361

The multi-choice system that we used in our first year game design course is shown in Table 2. The values balance the expected return from each category such that each is the best alternative in different situations. Category 1 spans 75-100%, category 2, 50-75% and category 3, 0-50%. This can be seen in Figure 2.

Teaching Balance

By exposing the process of balancing the options in the exam, we are able to give the students two perspectives on the same design principle. Further, we discussed in class the values to assign to category 2. Different values alter the strategy for answering. The discussion resulted in choosing 50% and 75% as the crossover points. This demonstrated the process of going from a desired balance to the individual values to achieve the goal.

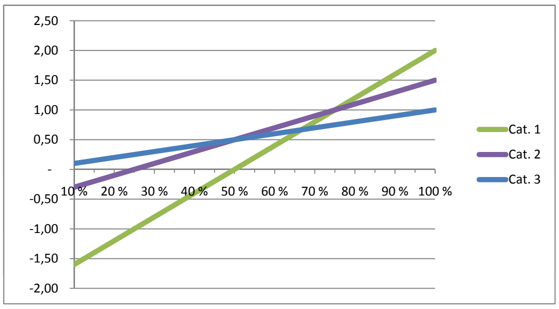


Figure 1: Graph showing the return for each marking category system. The highest point for any probability point (x) indicates the optimal category.

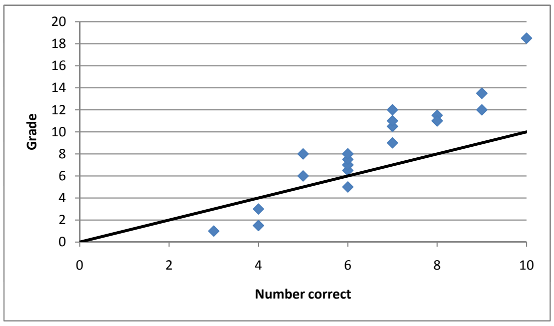


Figure 2: Final grade vs number of correct answers. Points above the line indicate students successfully using higher confidence values to increase their grade.

Results

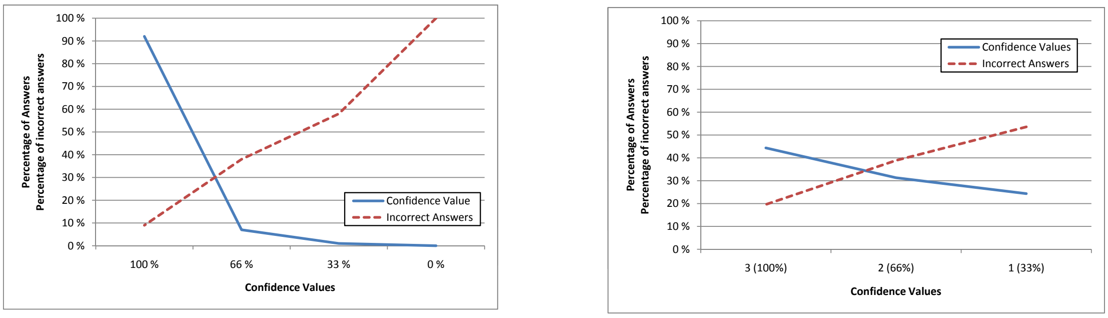
In a small class of 24 students, 23 sat the exam. The average number of correct answers was 66% with an average grade of 9.22 out of a maximum of 20 points. The distribution of confidence was 24% category 1, 31% category 2, and 47% for the highest level of confidence category 3.

As can be seen in Figure 2, students were generally able to increase their final grade by estimating their confidence. The black line indicates the grade achieved if the student selected the lowest level of confidence $(1, 0)$. The three lowest scores fall below this line as the students were over confident about their answers. These three students also had the lowest scores in the short answer section of the exam.

The results from the multi-choice section of the exam were positively correlated (r=0.79) with the short answer results. However, there was no significant difference in the correlation with or without confidence(r=0.79 and r=0.81 respectively).

Discussion

The response patterns from our students were different to those collected by Kolbitsch et al. [2008]. Students in the present study had far less confidence in their answers. This difference can be seen in the first data points in the TU Graz and our data in Figure 3. This difference can partly be explained by the design of our exam being intentionally difficult, to force the students to select levels of confidence other than 100%. Our data supports the conclusion that there is a positive correlation (r=0.4) between confidence and correctness of answer proposed in Kolbitsch et al.



A *B*

Figure 3: Comparison between the selection of confidence and the likelihood of providing an incorrect answer for the AOIS exam at TUGraz (*A*) (reprinted with permission), and IMT1361 exam (*B*)

Kolbitsch et al. were surprised by the number of students giving the highest level of confidence. However given our analysis of expected return, selecting the highest confidence level is the best strategy unless the student is less than 40% certain about their answer. In our results the spread of confidence is much wider, however it is still skewed toward higher confidence.

Perhaps the most interesting result is that the poorest performing students also have the least accurate assessment of their knowledge. This could partly be due to the nature of the distractors for many of the questions. These were designed to appeal to naïve subjects, but should not fool someone knowledgeable in the area.

Exposing the design process to the students and motivating understanding by using it as part of the exam seemed to have a beneficial effect on the students' learning. However, the students were not confident about their answers to the question about the marking system. Although this question was given by far the lowest confidence, it was answered correctly at nearly the average (61% vs average of 66%). This suggests that the students did not learn as much as we had hoped by exposing the marking system to them.

Student motivation and attendance in the first year design course was greater than that for the same group of students in other courses. We believe that by making the grading system more game like, it increased the motivation of the students to attend classes. It also provides an opportunity to discuss the design of courses outside the game development curriculum and general university assessment design.

The major divergence from game design in this experiment is the lack of feedback. Students do not get instant feedback during exams. In games the delay between action and result is generally very low. Without feedback on performance it is not possible to adapt your approach during the test. Therefore you cannot get better at the "game" (taking exams) until you associate the feedback with your choices. This is extremely difficult given the long delay between sitting an exam and receiving a grade, and the coarse granularity of the letter based grading system (A-F) used in Norway.

Case Study: XP in courses

The second case study we present is the use of role-playing game (RPG) character sheets and experience points (XP) as a way of tracking student performance through a degree. This section discusses the start of an experiment that will take several years to generate clear results.

RPG design and XP

One of the core features of role-playing games is the progression of the player character (PC). David Arneson introduced the use of experience points and character development as part of the game Dungeons and Dragons (D&D) in 1974 [Gygax and Arneson, 1974]. D&D has remained the most popular RPG since its introduction [C&GRmag, 2008].

The principle of using experience to indicate skill development has remained a core design element in both table top and digital RPGs. It is somewhat surprising to see the explicit representation of the XP system in computer games, such as World of Warcraft, as the system was initially designed as merely an accounting system, which could now be done by the computer. The numbers themselves are a form of feedback and remain popular even if the player does nothing with the information.

Designing the character development system is the core of creating the game mechanics in a RPG. Balancing rewards with effort and risk and providing multiple ways of progressing are important aspects of good RPGs. When a new challenge is too difficult for a player they often have two options - try to improve the use of their current abilities, or return to simpler activities to improve the character they are playing. The process of repeatedly returning to the same activity just to slowly gain experience points is called “grinding” [Koster, 2003]. This is perhaps the largest difference between the experience allocation system in these RPG games, and real world learning. Repeatedly performing a simple action is usually not enough to develop a new and different skill.

Instructional design and progress

Instructional design is a rich and active research area, where there are many pedagogical models. There does not however appear to be one unified theory of good instruction. There are, however, some generally agreed upon components of good instruction, including the need for effective feedback to the student [Shute, 2008].

Feedback can be broken down into two categories, short term and long term. Short term feedback is the immediate result of an action. This provides students with an indication of how well they are performing and, potentially, the direction in which improvements can be made. Long term feedback is the recording of progress over many learning sessions or modules. Short term feedback includes the immediate correction of answers in class, while the classic long term feedback is the school report or educational portfolio.

Short term, or immediate feedback can come directly from the environment, or be mediated through an instructor. Even when the environment is providing feedback, such as being unable to jump over a fence and falling to the ground, the feedback may still need interpretation to allow the student to improve on subsequent attempts. Long term feedback tends to be associated with social and cultural interaction.

Formative assessment is an important part of feedback when the goal is to assist the student in developing greater knowledge. This formative assessment must occur shortly after the exercise or it will lose much of its educational benefit [Boston, 2002]. Sweller et al. [1998] discusses the use of worked examples as feedback used to reduce cognitive load during learning. This is particularly important for multi-part problem solving.

Portfolios provide students with a longer term review of their progress, and act as a form of motivation and reflection. These portfolios track the significant achievements of the student over time, and can be used to monitor progress as well as provide examples of the student's previous success. Degree scale student portfolios are time consuming to maintain, and the benefit is not seen on an individual course level, so are not common.

One of the main difficulties with student maintained portfolios is motivation. Students are often relieved to have finished a exam rather than focus of internal reflection on their learning. Maintenance also requires translating our academic language of student outcomes and learning objects into something the student can understand and maintain.

Character Sheet development

As part of the Norwegian education system, all degree courses should be able to list a set of student outcomes that match to learning objectives. Documents such as the ACM curriculum documents [ACM, 2008], or the IDGA game development curriculum [IDGA, 2008], are often used as international standards. These documents, and the formal course descriptions built on them, use language and structures designed to meet the needs of administrators and review boards, not students. Course descriptions for students often use these academic documents as a foundation, and therefore do not engage students with language or metaphors they understand.

The use of an RPG style character sheet will hopefully increase students motivation to understand the learning objectives by providing them in a recognizable format.

The components of a classic character sheet are:

Description	Name, age, height, appearance, ...
Core Skills	Strength, intelligence, wisdom, constitution and charisma.
Equipment	Armour, weapons, magical clothing.
Learned skills	Weapon skills, fighting skills, adventuring skills, sundry skills.
Magical abilities	Spells, potions, scrolls.

We can alter these somewhat but the goal is to present the components of the degree in similar ways. The categories we have chosen match those above are:

Description	Home town, dialect / languages spoken, appearance.
Talents	Problem solving, learning style (assessed using the VARK scale), leadership, personality type (assessed using Myers-Briggs test).
Equipment	Computers, software, mobile phone, other electronics.
Learned skills	Programming, graphics, AI, OS, algorithms, math, game design, SE.
Networking	Social networking activities, company contacts, open source projects, competitions.

As students develop skills in particular areas they can record the level they have achieved. This provides a familiar model for the development of skills, and provides the student with an overview of both what is expected and what is possible.

The character sheet also provides an opportunity to recognize the secondary skills that are vital as part of a degree, but which do not fit into any particular subject area. Participating in the development community by entering competitions, working on open source projects, working on summer projects, or related disciplines are important parts of becoming a well rounded graduate. This is also very useful in situations where it is not possible to add a full course on a topic.

Building a team

When gamers form a group for a raid (a campaign in World of Warcraft), they are balancing player skills, time, and group cohesion, with character abilities. The ability to identify and utilize the skills in a team is vital for any group project. The character sheets provide a way that students to explicitly see the skills and build a team balancing these skills. Group projects can be created that require students who have focused on different areas of the curriculum, and can incorporate students at different levels from complementary courses. The ability to discuss ability openly is vital for effective scoping of projects, and character sheets provide a formal way of discussing ability.

The grades in assessed courses can be easily transferred into the level of the associated skills. The student must reflect on how well they performed in each area of the course, and assign a skill level to each. This provides a way to encourage reflection at the end of a course. Because this is calculated by the students it does not require input from the lecturers.

XP for student progress

Along with the character sheet, progress will be monitored using XP. This allows a unified number system between course work and extra curricular activities, and identifies that the skills learned as part of the curriculum are only a small part of the overall skills required of a graduate.

XP for course assessment

The use of experience points as the main means of assessing students' work within a course has recently been implemented by Lee Sheldon in the course T366: Multiplayer Game Design, at Indiana University, Bloomington [Sheldon, 2010a]. Sheldon treats the class itself as a game. The language used to describe classroom activities comes directly from massively multiplayer online games (MMOs). Assignments are quests, groups are guilds, exams are boss fights, and grades are expressed as levels achieved by gaining experience points.

Sheldon is an experienced game designer with many years of commercial MMO game design experience. The course is not merely using the language of games but is designed from a game designers perspective. The use of experience points as a measure of progress is critical to the structure of the course. Each activity is allocated a number of experience points and the final grade is mapped from the level achieved by gaining XP. Jesse Shell describes the XP based approach as similar to normal assessment and grading with the difference being [Schell]:

- Averages are delayed.
- Grades monotonically increase with XP earned.
- Familiar and fun associated with the name XP.
- Easy addition of extra credit.

XP for extra curricular activities

One of the difficulties when designing a complete study program is to ensure that the courses within the degree fully cover the desired graduate profile. Often there are skills or attributes which do not fit into the learning outcomes of any particular course. These skills are often spread around courses or accumulated into fragmented "advanced skills" courses.

Using the game design approach of allocating XP allows us to recognize development in these extra curricular areas. The graduate profile for our game programming degree includes attributes such as: knowledge of current games, the ability to discuss current trends in game technology, an awareness of the cultural and social impact of games, and engagement in the gaming community. These attributes do not clearly reside within any of the technical courses which we teach. Some cannot be assessed within a single semester, or assessed as part of regular courses.

We will be implementing an XP based system in two ways in the fall 2010 semester. In the first lecture, students will fill out an "adventure" character sheet describing their current skills and background. At regular intervals, and at the end of each semester, this character sheet will be updated to reflect the progress of the student through the degree. From the instructional design perspective this is equivalent to the creation of a portfolio that will follow the student through their entire degree.

Students will also be provided with opportunities to earn additional XP by completing extra curricular activities. These focus on exploring the subject "world" in depth, and reward students who strive to learn material beyond the standard course material.

Having earned enough XP, the student will "level up" which will provide various benefits and skill points to spend. These can be spent on various course activities, including adjusting the weighting within assignments, additional tutoring, and access to research equipment such as VR goggles, brain-computer interfaces, augmented reality drones, and mobile devices. This provides rewards that are within the learning objectives of the degree program, but would not be possible to provide for every student.

Discussion

In the postmortem for Sheldon's MMO course the, students identified many features of the course which they enjoyed but also several aspects that could be improved. Students enjoyed the opportunity to "game the course" [Sheldon, 2010b], and that the course used language that was familiar to them. They were, however, disappointed that not all activities were given XP points. Having built an expectation that every action would be coupled with reward, those activities that did not have credit associated with them were seen as inconsistent and irritating.

A further problem was the balance between public and private information.

There is tension between using public high score lists and recognition of leveling, with the right to privacy. Most of the students prefer to use an alias rather than their real names when communicating online. This has recently been shown by the controversy over Blizzard's proposal to use real names in the World of Warcraft and Starcraft2 forums. Player anger resulted in a backdown from Blizzard and a return to using aliases. Lecturers must be cautious when publicising grades and achievements.

This paper has discussed the use of game design and game metaphors in the design and delivery of tertiary curricula. We feel that the using game design principles to help design our courses improve their alignment with the students background and experiences. There is alway more to learn, and given that lecturer enthusiasm is a significant contributing factor in student learning, we should strive to find teaching styles that stimulate the enthusiasm of the lecturer as much as the students. As the gamer generation becomes academics we can expect to see an increase in game like systems being used as the basis for teaching.

References

ACM. Computer Science Curriculum 2008: An Interim Revision of CS 2001. Technical report, ACM, December 2008.

AndroLib. Androlib.com. http://www.androlib.com/appstats.aspx, 2010. [Online; accessed 23-05-2010].

Carol Boston. The concept of formative assessement. *Practical Assessment, Research & Evaluation*, 8(9), 2002.

C&GRmag. Comics & Games Retailer magazine. http://www.therpgsite.com/showthread.php?t=6959&page=8, February 2008. Issue 191.

Malcolm Walter Corney, Donna M. Teague, and Richard N. Thomas. Engaging students in programming. In *12th Australasian Computing Education Conference (ACE 2010)*, pages 18–21, Brisbane, January 2010.

Mihàly Csìkszentmihàlyi. *Flow : the psychology of optimal experience*. Harper & Row, New York, 1st edition, 1990. ISBN 0060162538.

FADE. Forcasting and Analysing Digital Entertainment. http://fadellc.com/Press.html, 2010. [Online; accessed 23-05-2010].

Gudmund S. Frandsen and Michael I. Schwartzbach. A singular choice for multiple choice. *SIGCSE Bulletin*, 38(4):34–38, 2006. ISSN 0097-8418.

James Paul Gee. *What Video Games Have to Teach Us About Learning and Literacy*. Palgrave Macmillan, 2004. ISBN 1403965382.

Gary Gygax and Dave Arneson. Dundeons and Dragons. TSR, 1974.

IDGA. Curriculum framework, v3.2beta. url-http://wiki.igda.org/images/e/ee/Igda2008cf.pdf, Feburary 2008.

Mary J. Irwin. Cooking up a blockbuster game, forbes.com. http://www.forbes.com/2008/11/21/games-eedar-developers-tech-ebiz-cx_mji_1121eedar.html, November 21 2008. [Online; accessed 23-05-2010].

J. Kolbitsch, M. Ebner, W. Nagler, and N. Scerbakov. Can confidence assessment enhance traditional multiple-choice testing? In *Interactive Computer Aided Learning, ICL2008*, 2008. Villach.

Raph Koster. Small worlds. In *Game Developers Conference*, 2003.

M. Prensky. *Don't bother me Mom-I'm learning*. Paragon House Publishers, Minneapolis, 2007.

Marc Prensky. Digital natives, digital immigrants. *On the Horizon*, 9(5), October 2001. MCB University Press.

Anthony Robins. Learning edge momentum: a new account of outcomes in cs1. *Computer Science Education*, 20(1):37–71, 2010. URL http://www.informaworld.com/10.1080/08993401003612167.

Jesse Schell. Lee Sheldon gets slashdotted. http://gamepocalypsenow.blogspot.com/2010/03/lee-sheldon-gets-slashdotted.html. [Online; accessed 23-05-2010].

Lee Sheldon. T366: Multiplayer game design. http://gamingtheclassroom.wordpress.com/, 2010a. Syllabus [Online; accessed 23-05-2010].

Lee Sheldon. T366: Post mortem. http://gamingtheclassroom.wordpress.com/, 2010b. Post mortem [Online; accessed 23-05-2010].

Valerie J. Shute. Focus on Formative Feedback. *Review of Educational Research*, 78(1):153–189, 2008. URL http://rer.sagepub.com/cgi/content/abstract/78/1/153.

Statistics Norway. Norwegian media barometer. http://www.ssb.no/medie_en/. Tables in StatBank [Online; accessed 23-05-2010].

John Sweller, Jeroen J. G. van Merrienboer, and Fred G. W. C. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251 – 296, September 1998.