

Formative Research for Game Design

Jay Bachhuber

Center for Children & Technology/Education Development Center, Inc.

96 Morton Street, 7th floor, New York, NY 10014

November, 2012

For more information,

visit <http://cct.edc.org> or <http://possibleworlds.edc.org>

Preferred citation: Bachhuber, J. (2012). *Formative research for game design*. New York, NY: Center for Children & Technology/Education Development Center, Inc. Retrieved from <http://possibleworlds.edc.org>.

Formative Research for Game Design

Jay Bachhuber

Center For Children & Technology/Education Development Center, Inc.

November, 2012

Educational game designers fight an uphill battle. The audience they target—academically struggling students—is the least likely population to seek out games designed around academic skills or content. Students having difficulties with a subject or skill set are not going to look to games addressing those topics for fun or entertainment. When approaching this challenge, educational game developers often fall into one of two camps. Designers tend to craft their games from hunches and insights gained from personal experience, while academics build their games around accepted psychological or pedagogical theories.

Each approach has its merits but, too often, each approach falls short, producing either entertaining but thinly educational games, or pedagogically sound but dull ones. By focusing on the skills that struggling students need to develop, rather than the skills they have now, educational games can alienate the very population for whom they're designed. I don't mean to criticize these game developers, because designing a game for the specific skills and mental habits of struggling students is an incredibly challenging task. This is true in part because, while research that explains what children do not know how to do is common, much less research on learning investigates how children approach academic tasks they do not yet understand, or how they attempt to reason about material they have not yet been taught. This gap in the research presents an opportunity for developmental psychologists to collaborate with game designers and greatly improve the quality of educational games.

During our development work at CCT, we've found formative research to be an invaluable tool because it allows us to gain insight into how our target audience perceives and responds to learning goals they have not yet achieved. These insights allow us to create designs that meet students where they are and engage them in game play that helps them build new knowledge. By discussing concrete examples from the design process of one of our projects, this chapter is intended to acquaint readers with the practices and benefits of this practical methodology.

What formative research is

During the game-development process, there often are design questions that have no clear answers, because psychological development theories could support a number of options and so have limited predictive power for the efficacy of specific design choices. Formative research is a way to resolve these issues by using appropriate stimuli to probe how children make sense of specific tasks or media.

The work is not meant to expand what is known about how students in general understand a given topic, nor is it about assessing the knowledge of any specific child. Instead, research participants are viewed as members of a particular subgroup, and

developmental psychology provides theoretical models against which student responses can be checked. This student category should strike a balance between being broad enough to be useful in diverse classrooms, and narrow enough that, during formative testing, confusion for any individual student or similar responses among a few is sufficient to inform design decisions.

Methodologically similar to clinical interviewing as practiced by Jean Piaget, and more recently by Herbert Ginsburg, and others, formative research is fruitful when interviewees feel comfortable speaking honestly without concern for getting answers right or wrong. If they're relaxed and engaged with the task, it's also likely that they're going to speak imprecisely, muddling through without carefully choosing their words. The researcher then must be able to think along with the student, asking her to clarify without giving any sense of judgment, while noticing the salient actions, comments, or even single words that should be further explored. I've found this process to be very similar to that of interviewing sources for newspaper articles, and my background as a journalist has served me well here.

The researcher should continually open up the space for new ideas without using leading questions, and invite the student to clarify or give more detail without feeling like her previous answers were wrong or inadequate. Finally, when the interviewer has either heard enough to serve the purpose of the interview, or the interviewee appears to be feeling frustrated or stuck, the interviewer moves on to a new question or ends the interview. When practiced well, formative research can be a reliable way of ensuring the artifact being designed has the best chance to accomplish its purpose.

What the process looks like

Formative research can look slightly different, depending on whether it's occurring before or after the creation of a prototype, though the process and goals are fundamentally the same in either instance. In the first situation, evocative objects and activities are used to prompt a child's thinking about a specific topic, and we often create testing materials that don't resemble the eventual, final product. Instead, the designed activity is meant to target student thinking in specific contexts. Gaining insight into the student's thought process can eventually help us design better games.

Once preliminary testing has established the instructional design, a prototype of the educational media, or its current iteration, is used to help designers understand if it is eliciting the intended type of thinking. Here, existing theory has informed certain design choices but research is required to observe whether children make sense of it as expected, or how the media provokes unintended thinking and what that thinking is.

While formative research doesn't replace the experience and intuition of designers, it provides useful guidance and constraints. As long as the interviewer has a clear idea of the information he's looking for, and the ability to improvise while staying true to a disciplined methodology, this research can be done with minimal preparation, relying on only small groups of children. Formative research can be planned and implemented within a day or two, answering design questions as they arise, in order to keep the development process moving forward.

Using formative research to guide development

In 2009, EDC|CCT received funding from the Bill & Melinda Gates Foundation for the Portable Wordplay project, a two-year research and development effort with two primary goals. First, the project explored the feasibility of designing instructionally rich digital games that could engage seventh graders in sustained, playful literacy practices relevant to improving their reading comprehension skills. Second, the project demonstrated that digital reading comprehension games can be integrated with regular classroom instruction in ways that are both logistically manageable for teachers and beneficial to student learning.

The project was intended to respond to the academic needs of students who have basic decoding skills, but lack adequate fluency and comprehension strategies and struggle with understanding the more complex texts required of them during middle school. A lack of vocabulary knowledge can compound reading comprehension challenges and inhibit students from engaging with content that they are cognitively ready to handle. We decided that it could be valuable to make a game focused not on specific vocabulary words, but instead on vocabulary skills that can improve comprehension.

Digital games offer a number of unique affordances that we've tried to leverage in our designs. Games offer unthreatening play spaces in which students can construct and evaluate new conceptual knowledge, adapt to changing rule sets, and experiment with unfamiliar ideas while pursuing motivating goals. In well-designed games, players pay close attention to the environment and their actions within that environment, looking for relationships between their actions and desired outcomes.

With these ideas in mind, we developed and field-tested two related Nintendo DSi games—*Code Invaders* and *Cipher Force*—with supporting materials to help teachers integrate the games with traditional literacy instruction. Based on observations from the test and on teacher feedback, we translated the games into an asynchronous Web experience and tested that in order to better pursue our two principal project goals.

Research-led development

In spring 2010, we had established broad goals for the Portable Wordplay project and we soon decided to focus on multiple-meaning words. Four principles underlay our design decisions: (1) word play is beneficial to student literacy learning, and a game that motivates students to spend time with words can have positive effects beyond those measured in standardized tests; (2) students with basic, below-grade-level literacy skills still can be capable of other complex challenges, and the game should take advantage of this to be challenging and fun for a wide audience; (3) the game must not privilege skilled readers over struggling readers; and (4) the game by itself does not need to teach vocabulary, but rather to support traditional classroom instruction in a blended model.

We knew that, when reading complex texts, students sometimes interpret the wrong definition of a multiple-meaning word, and, without the metacognitive habit of monitoring comprehension while reading, may not notice that they've lost track of a

text's logic. A game that incorporated the strategies that strong readers use to disambiguate word meaning seemed like a useful tool for struggling comprehenders.

While we knew the reading strategies that are taught to seventh graders, we found that the literature generally describes what struggling readers *should* do when reading challenging texts, but offers very little about what struggling readers *actually* do. This is precisely the challenge formative research can address to help bring together the perspectives of designers and academics. We needed to understand the thinking of struggling readers in order to design a game that would be comprehensible to them and would help them move towards a richer relationship with text, where they understand words not as static blocks of meaning but as flexible tools that serve communication.

Stage 1: Speakers and multiple meanings

One idea we initially considered was a game design involving communities of usage. We knew that in different fields, such as architecture or computer programming, the same words can have different meanings. New words are often brought into usage and used in unique ways to solve particular communication challenges. When encountering an unknown word, considering the community of usage is one strategy for narrowing down a word's possible meanings. We thought that communities of usage could be a way to make the sometimes subtle differences between word meanings more concrete for students. Different-looking speakers could create a salient cue for students to consider alternate interpretations of a word in a sentence. To understand if this was a reasonable approach, we first wanted to hear how students think about what they know about words, and how they discern and interpret word connotations.

Our initial formative testing activity involved pictures of different speakers, ambiguous phrases such as "I live on a higher plane ..." and a number of potential phrases to finish the statement. Students were asked to match a photo of a speaker with the ending phrase that the student thought made the most sense for that speaker. For example, students were shown the phrase "I live on a higher plane ..." with a picture of a meditating guru. They were asked to choose an appropriate ending to the sentence from phrases including "now that I have risen above concerns of the flesh," "up on that plateau in the mountains," and "enjoying luxuries I never knew existed." I designed the materials such that the six speakers and their quotes corresponded to six different usage categories of the phrase "higher plane," including status, spiritual, and literal.

While we intended the sentences and images to fit together in a logical way, our interest was not whether the students would match the ending phrases with the speakers as we had. Instead, what was most interesting was their process - how they thought through the different possible meanings of "higher" and "plane," and what their reasoning was for pairing the meaning of the sentences, as they understood them, with the photos of the speakers, as they understood them.

To conduct the research, I went to an unstructured afterschool program where students have free time to play on computers. I met individually with four students, explaining to each that there were no right or wrong answers; I was simply interested in how they thought through the activity. It's important to emphasize this point before working with students, and to keep tone and body language casual during the interview, because students expect to be quizzed by adults and can be shy if they feel they can't provide the

adult with what they think is expected of them. It can take a while before they understand and accept that this is a different type of interaction, where I am genuinely interested in their thinking regardless of what either of us knows about the topic. The tone of these interviews was meant to be, as I was taught, “like bar talk.” It was important the students felt we were considering the activity together out of a simple, inconsequential, and unhurried curiosity.

While working on this activity with students, I saw them interpret the content in a wide range of ways, but the processes by which they interpreted the activity shared many commonalities. Part of what I saw was that students were frequently unfamiliar with the speakers used in the exercise, and interpreted them very differently from how we had imagined them. For example, one student thought a picture of a pilot was a picture of a bus driver, and so this speaker had no clear connection to any of the sentences. The speaker-as-scaffolding approach seemed to be of no help because students had such different background knowledge.

Even when students did pair speakers and ending phrases as I had intended, I was not sure their interpretations of “higher plane” were the same as mine. It was entirely possible that a student had completed the activity as I’d intended purely by chance, or using a set of assumptions and decisions that were not at all the assumptions and decisions I’d imagined when designing the activity. In these situations, asking non-judgmental or non-evaluative questions like, “How do you know that?” or “How did you decide that?” can be useful for uncovering the real thinking of a student. As I’ll explain, a variation on the activity can also be developed on the fly.

Having seen students interpret the speakers and sentences themselves, I was interested to see how they would explain alternate pairings. I matched speakers and ending phrases as I understood them, and asked students what they thought “higher plane” might mean as I’d arranged the materials. Doing this, they produced very creative, if convoluted, explanations of the sentence meanings. Rather than consider alternate connotations of “higher plane” to fit the speaker identity and ending phrase, they held onto their initial understanding of “higher plane” and bent the sentences to fit it.

It’s common for an interviewee to describe what she’s doing or thinking one way, while the researcher notices that something else is actually going on. This happens, in part, because, as mentioned earlier, young people are used to relating to adults in ways that accommodate adult expectations. Students may think they know what the “right” answer is in an activity (“right” in that it’s what the researcher wants/expects to hear), and may give that answer even if it doesn’t correspond to their real feelings or understanding. In these situations, the researcher can try to open up the discussion to see if the child responds. By continually providing space for new ideas, the researcher can check the accuracy of how he is understanding the child’s experience.

Additionally, when conducting formative research, it’s important to have evidence criteria to evaluate whether one is accurately interpreting what one is seeing and hearing. It’s important to listen to a student’s explanation, consider what evidence would confirm that the student is actually doing what she claims or what the research thinks, and then ask questions and prompt activity that should generate this evidence.

After asking students to explain their thinking during the exercise, I also asked more general questions about their knowledge of word meanings and what they do when encountering unknown words and phrases while reading. Most of the students I worked with said that they used context clues to figure out meaning when they didn't know it, and this commonality seemed to warrant further exploration. I wasn't sure if students had simply been told so often by their teachers to "use context clues" that they thought this was the answer I wanted to hear, or if they actually did use the strategy but poor understanding of word meanings led them to misinterpret the surrounding phrases that could create context. During the "higher plane" activity, struggling readers didn't seem to consider that words could have alternate meanings they did not know, and I thought this could impair their ability to effectively use context to understand unknown words.

Stage two: Images and multiple meanings

For the next formative activity, we decided to stop pursuing images of different speakers as scaffolding. Instead, we were interested in seeing if other images could act as scaffolding to help students discern the meaning of multiple-meaning words in an ambiguous phrase or sentence. We thought that by watching students interpret unknown words by using image-based context clues instead of textual ones, we could better understand their thinking unhindered by the burden of decoding.

To begin, we presented the word "complex" in isolation and asked students to talk about anything they knew about that word. Next, we showed the phrase "complex survey" and asked what they thought that phrase might mean. We added two pictures related to the phrase "complex survey"—a complicated-looking subway map and a survey with check boxes—and asked, "What do you think about 'complex survey' when you see these pictures?" After discussing this, we kept the pictures, but replaced the phrase "complex survey" with "It was a complex survey. There were many parts and sections to fill out," and asked, "What would you think 'complex survey' meant here?" Finally, we removed the pictures and sentences and repeated the process with "It was a complex survey. The department of housing wanted to see how many blocks the buildings covered" and pictures of a person surveying land and a building complex.

From this and a following testing session, we gained a number of insights and one key question that would determine the game's ultimate focus. While students struggled to understand unknown words, and were hesitant to express ideas about what an unknown word in context might mean, they had no such reservations when looking at a picture that was unfamiliar to them and offering interpretations of it. Also, images could jar students, and make them question what they knew about a word. This was particularly exciting. One girl initially explained that complex means difficult, and that a survey is a group of questions. A complex survey, then, was a survey that was difficult to understand. When I showed the sentence "It was a complex survey. There were many parts and sections to fill out," she readily explained that the survey was hard to fill out because of all the questions. When shown the picture of the survey form and the many interconnected lines of the subway map, however, she was confused. The survey photo fit with her understanding of survey, but she could not connect her understanding of complex with the subway map. This confusion provided an opportunity to begin a conversation about

the different definitions of complex, and how something complex can be hard to understand (as the student understood it to mean) but that this was not really synonymous with “complicated.” While the pictures were not sufficient scaffolding to help a student understand an unknown word, or unknown usage of a word in a context, the images were helpful to begin a discussion about a word meaning.

If struggling comprehenders have a tendency to use word meanings that are inappropriate or insufficient, causing them to misinterpret sentences, then a cue that makes them reconsider their interpretation of a word would be very useful. While we had imagined that pictures might be a way to provide context to understand unknown word meanings, what we found was that perhaps the opposite would work better: New word meanings could help students understand confusing pictures and, in the process, come to better understand a word. Because we were designing the game to work in a blended model with traditional instruction, we reasoned that teachers could build on the work students did in the game and relate those interpretation skills to traditional use of in-text context clues.

During this same session of formative testing, we also had an activity in which a student saw three sentences containing a multiple-meaning word, one sentence at a time, each one having more detail than the one before. Each sentence was accompanied by three words, and I explained that one of them was the meaning of the unknown word. One group of sentences used the word “excruciating,” and one of the options for its meaning was “mild.” A student I was working with read the first sentence, and after reading the word mild confidently expressed that mild was not the answer because he knew what that meant. When asked to explain, he paused for a moment and said, “I know mild chicken,” and explained that mild chicken is close to spicy, but not spicy, so mild means “close to something but not it.” He knew the word mild from the context of a fast food restaurant and developed his own quite reasonable and functional, though inaccurate, definition.

During instances like this, it was important for us as researchers to look without judging, a perspective that underlies all formative research. At the time, it was not important what the student thought the definition of mild was, it was important *why* he thought it and how he came to think it. After the student explained how he “knew mild chicken,” my follow-up questions revolved around how he came to know other words, and how, when encountering a word that was unfamiliar or only slightly familiar, he pieced together potential meanings for it.

From the literature, we knew that vocabulary learning is not all-or-nothing, but rather is incremental. People can learn more about a word each time it is encountered. From the mild chicken conversation, and others, we saw the other side of this incremental language-learning process: If never checked against dictionary definitions, misconceptions about a word also can be continually reinforced from repeated exposures.

After this round of testing, a key question about student thinking remained. We were still uncertain of exactly how students understood what word definitions are. It was unclear whether students thought of words as having distinct, sometimes unrelated, meanings that could be used in a variety of ways, or whether they thought of words as having “fuzzy” meanings that overlapped and could be used in virtually limitless ways. Continued testing focused on resolving this question.

Building on formative feedback

In the two months that followed our initial testing sessions, we tried a range of materials and activities to help us better understand how students understood word meanings. We came to label the types of word knowledge we were seeing with the very sophisticated terms “mush ball” and “rough hewn.” In the first case, the student had likely encountered a number of a word’s meanings and was not differentiating among them. A word could have a number of meanings, possibly even contradictory ones, that were all rolled together. This was especially true of words with subtle distinctions such as “set.” A rough-hewn understanding was similar to the dictionary definition of a word, but still quite vague. This is more like the student who explained that the word “complex” means “hard” or “difficult.” While in many common usages complexity can imply difficulty (“a complex problem”), difficulty is only one part of one definition of complex.

From the beginning of the project, we’d intended to address vocabulary concepts more than specific vocabulary words, and our formative work seemed to bear out the utility of this approach. Misconceptions about word definitions appeared to be pervasive among struggling readers, and acted as real barriers to literacy proficiency. These months of testing also helped us resolve a question about which level of word understanding our game should prioritize. We had considered three options: language in any community at any register, e.g., science classroom, a family dinner, school playground; multiple communities within one register, e.g., different classes at school; and, finally, just one community in one register, e.g., an ELA class.

Reflecting on what I saw during this research, our team decided to stop exploring communities of usage as a form of scaffolding because identifying a speaker and using that understanding to decipher a sentence required background knowledge that we couldn’t rely on all students to have. We also decided that difference in word connotation when used by different communities was too sophisticated a concept for struggling comprehenders. We’d thought that students were more likely to understand a multiple-meaning word’s different connotations in different situations than its distinct, formal definitions. Instead, we learned that the differences between how different communities use a word are often subtle, and that accurately interpreting them requires first understanding the formal definition of a word.

Though different communities may be more likely to use one of a word’s definitions (for example, a band teacher and gym teacher are likely to use different definitions of the word score), this disambiguation technique didn’t seem worth including in our game. This freed us to focus on helping students better understand word meanings within one register, such as school. We also decided it was important for the game to encourage students to attend to the specific, distinct definitions of multiple-meaning words. One way to do this, we thought, could be to have students use images to communicate word definitions to each other.

From both the literature and our formative work, we understood that adolescents are particularly interested in communication. They’re beginning to understand the world in more complex ways, to be able to discern contradictions and inconsistencies, and to place greater importance on socializing with each other. Developmentally, they’re ready to

consider alternate points of view and to imagine the perspectives and thinking of other people. It seemed likely that we could leverage this in our game design, encouraging players to consider word meaning through a social game with mechanics emphasizing clear communication. Using pictures instead of text to communicate would hopefully put struggling and confident readers on more equal footing.

After brainstorming game ideas, we created a paper prototype in which students alternated playing two roles. The *codemaster* was given a word with three definitions and tried to communicate one of the word's definition using three images, called an image code. The *guesser* tried to interpret the image code to choose which definition from five possible words was being depicted. We took several early versions of this activity out to test with kids.

From testing, we learned three things fairly quickly. First, this activity was kind of fun, and there was real potential to make an engaging game. Using pictures to communicate did help readers with varying proficiencies play together, and they liked trying to figure out each other's thinking. None of images available to the codemaster depicted exactly what her definitions said, and oftentimes students found this frustrating initially. But after a few rounds they came to appreciate the opportunity this provided for creative play. This gave us the encouragement to stick with the game's core mechanic of making and guessing codes, and to begin refining the game design.

Second, we saw that showing the codemaster all three definitions and having the guesser guess the word undermined the game's purpose. We wanted the activity to help the codemaster see the distinctions among a word's different definitions. In practice, however, codemasters would often make an image code referencing all three definitions, either picking an image for each definition instead of using three pictures for one meaning, or combining the definitions in their head and making a code for this new amalgam. Guessers also weren't attending to the specific definitions, instead focusing on the words from which they were choosing. We adjusted the game so that the codemaster was shown only one definition, and the guesser had to choose among a minimum of three definitions for one word or a maximum of nine total definitions among three words.

Third, if the players felt they were competing, they would try to trick each other by creating image codes that associated with the chosen definition only through extremely convoluted or idiosyncratic logic. This was an interesting challenge: how to get kids excited about cooperating, while keeping the motivation that some kids feel from competition? Throughout all our testing, we continued to see indications that our core activity worked pedagogically as we intended, but making it function as a game took a significant amount of research, revision, and work with our production partners.

After two field tests and a number of iterations, we now have a version of the game that we're really proud of. In the current version, play now occurs asynchronously at home instead of together during class. While the current version's gameplay is quite a bit different from our original game design, the dynamic of students making and guessing codes using images has remained, and has been refined to allow easier integration into traditional classrooms, play between students of different abilities, and player focus on understanding and communicating the distinct definitions of multiple-meaning words.

Now teachers also can customize a game for their class by choosing the vocabulary words students will be playing with, how many words and definitions the players will be choosing from when they're guessing a teammate's code, and how many points the class is collectively trying to achieve. Once assigned a game, students go online and log into their account on the game site. They then enter the game created by their teacher, and choose from a library of photos to create image codes for different word definitions. Each code consists of three pictures, and can include custom images the students make using geometric shapes. Players also can alter the pictures they choose, using arrows and cropping to draw attention to or away from elements of the picture.

The following night, players log in again and try to guess the image codes created by their classmates. Teachers can view all the students' codes and reference them later while facilitating conversations about a variety of literacy topics.

Future Steps

The field tests that have helped us improve our game design and implementation model also have provided opportunities for us to check whether core features of the games that we focused on in formative testing are in evidence when kids are playing the games alone. As noted earlier in this chapter, three of our four design principles were that (1) word play can help students improve their literacy skills and, therefore, (2) we should make a game that is challenging and fun for students (3) without privileging literacy skills. Having seen students of different reading abilities play together well, we're optimistic that *Cipher Force* is holding true to those principles. We've heard from teachers that students who normally don't participate much in their classes have been eager to contribute, and even take on mentor roles with students less familiar with gaming conventions. We're excited to see that our games seem to encourage struggling readers to enjoy word games.

Our fourth design principle was to make a game that integrated well with current classroom practices and helped teachers. Moving the games from the Nintendo DSi to the Internet has helped with that, and we're now launching our games on the educational media aggregation site Edmodo. With the professional development videos we're offering with the games, we think they'll be easier than ever for teachers and students to use.

As researchers like Donna Alvermann have noted, students who perform below accepted norms on tasks related to academic texts often can show considerable literacy proficiency when working with other types of texts. The problem is not always with the students and their abilities, but rather with how their abilities in specific areas are prioritized or invalidated. With formative researchers serving as advocates for student abilities, however, I hope more educational projects can be designed to help students as they are, rather than as we'd like them to be.