Multimodal Learning Through Media: What the Research Says

By Metiri Group – Commissioned by Cisco

Contacts:

Charles Fadel, Global Lead, Education; Cisco Systems, Inc.: cfadel@cisco.com

Cheryl Lemke, CEO, Metiri Group: clemke@metiri.com

Foreword

This report is the third in a series that addresses "what the research says," as Education's ethos states that "without data you are just another opinion." Like its widely quoted predecessor, the "Technology in Schools" report (www.cisco.com/web/strategy/docs/education/TechnologyinSchoolsReport.pdf), and the newly introduced "Education and Economic Growth" (http://www.cisco.com/web/strategy/docs/education/Education-and-Economic-Growth.pdf), this report intends to provide a grounding in facts that can benefit the entire Education arena, from pre-K–12 to higher education, corporate training and development, and lifelong learning.

There is a lot of misinformation circulating about the effectiveness of multimodal learning, some of it seemingly fabricated for convenience. As curriculum designers embrace multimedia and technology wholeheartedly, we considered it important to set the record straight, in the interest of the most effective teaching and learning.

As always, your welcome feedback will allow us improve the report, or suggest avenues for future papers.

Happy reading,

Charles Fadel,
Global Lead, Education
Cisco

Introduction

> "A picture is worth a thousand words."
> –Author Unknown

People have long quoted this statement, often attributing it to an ancient Chinese proverb. Emergent neuroscience and visualization research now reveals glimpses of the science behind the saying. Visuals matter. The rapid advances of technology in literally every field, including communication, medicine, transportation, agriculture, biotechnology, aerospace, and energy, have tremendously increased the amount of data and information at our fingertips. As we strive to make sense of unimaginably large volumes of data, visualization has become increasingly important. Why? Our brains are wired to process visual input very differently from text, audio, and sound. Recent technological advances through functional Magnetic Resonance Imaging (fMRI) scans confirm a dual coding system through which visuals and text/auditory input are processed in separate channels, presenting the potential for simultaneous augmentation of learning. The bottom line is that students using well-designed combinations of visuals and text learn more than students who only use text.

A Myth Shattered: Bogus Data

Educators are in constant search for more efficient and effective ways to advance student learning. Thus it is no surprise that educators have been interested in the often-quoted saying that:

> **We remember…**
>
> 10% of what we read
> 20% of what we hear
> 30% of what we see
> 50% of what we see and hear
> 70% of what we say
> 90% of what we say and do

Unfortunately, these oft-quoted statistics are unsubstantiated. If most educators stopped to consider the percentages, they would ask serious questions about the citation. They would inquire about the suspicious rounding of the percentages to multiples of ten, and the unlikelihood that learners would remember 90 percent of anything, regardless of the learning approach.

Despite these obvious signals, many people have blindly perpetuated these mythical statistics without ever checking the source. Following are just a few of the many examples where this data has been inappropriately used. (Because all instances could not be included, the specific citations used as examples here are not referenced.) Readers should conduct a Web search with the term "cone of learning" or "10% of what we read" to see firsthand the extent to which these incorrect statistics are perpetuated.

Figure 1. **Cited by a U.S. Company**
This graphic was accessed from the Website of an eLearning company.

The source cited by the company is Edgar Dale's *Audio-Visual Methods in Technology*, Holt, Rinehart and Winston.

If representatives from the company had researched the actual text of the citation (which is out of print but still accessible), they would have found that Edgar Dale's visual did not include percentages.

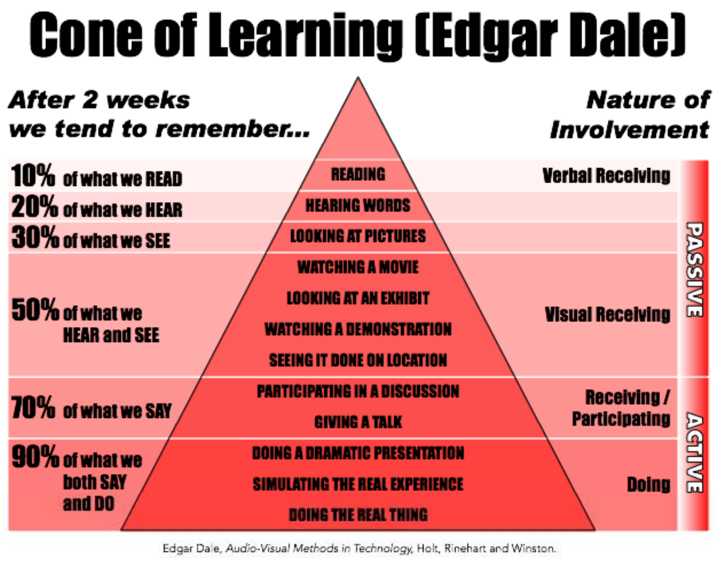


Figure 2. **Cited by a Major U.S. University**
This graphic was accessed from the Website of a major U.S. public university.

The source stated that the graphic was "developed and revised by Bruce Hyland from material by Edgar Dale."

Unfortunately, the site does not provide a citation for the work by Bruce Hyland, instead citing Edgar Dale's book, which, as mentioned earlier, does not include the statistics.

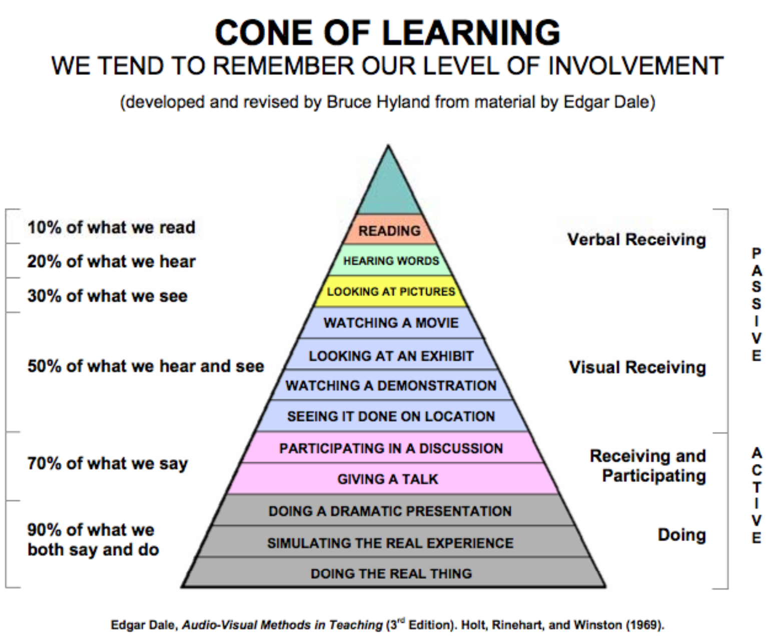


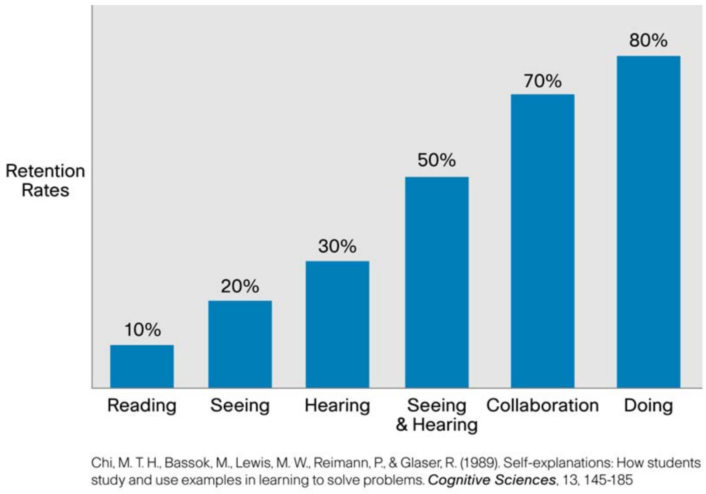
Figure 3. An Adaptation by a Private University
This graphic was accessed from a private university Website.

The site shows a derivative of Edgar Dale's cone and then establishes the Learning Pyramid. That pyramid includes average retention rates. The only reference is, "The National Training Laboratories in Bethel, Maine," No other citation is provided.

Figure 4. Cited in a Presentation by an e-Learning Company

This graph was accessed through personal correspondence with a representative of a technology firm.

The source cited (e.g., Chi et al., 1989) does not contain the referenced graph.

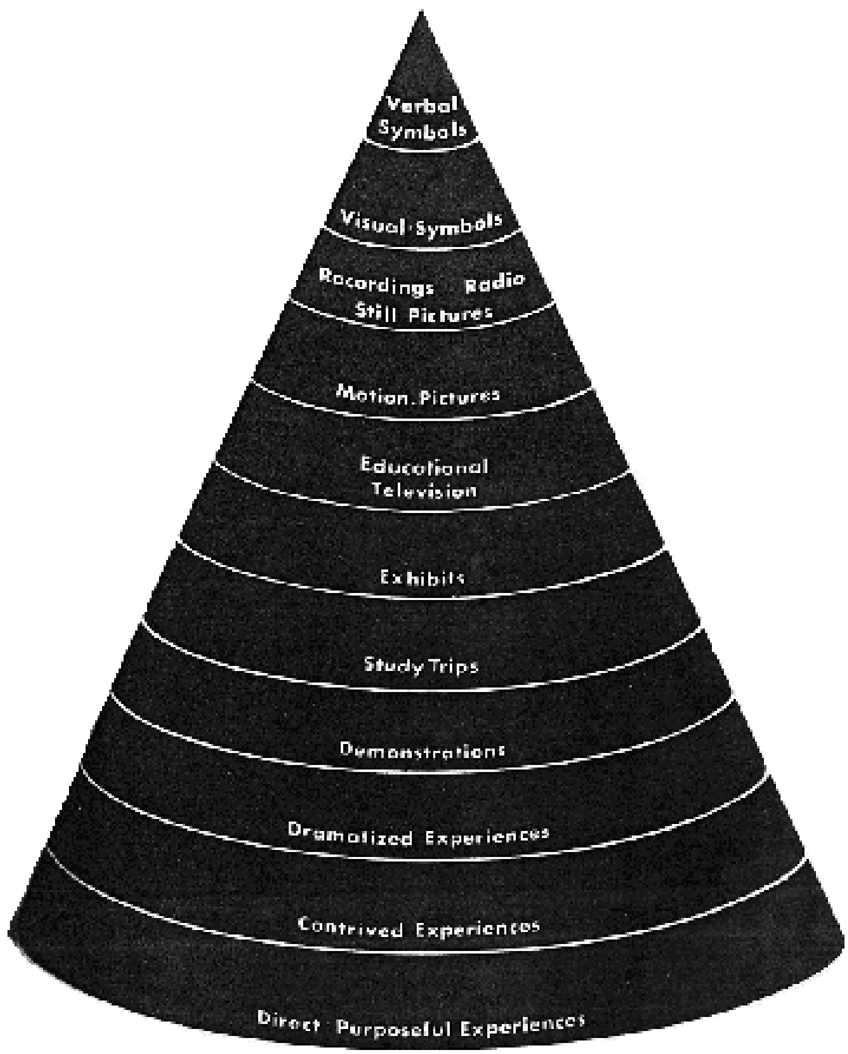


Tracing the History of the Myth

Edgar Dale (1954), an early researcher in the field of visual learning and the father of the Cone of Experience, is credited for the original linkage between instructional theory and communications media. Unfortunately, he is inaccurately credited with conducting the research behind the bogus "data" associated with his cone. In fact, Dale's original model of the cone does not include any percentages, and is explicitly described by Dale as a visual aid about audio-visual materials. Dale's cone of experience is essentially a "visual metaphor" depicting types of learning, from the concrete to the abstract. Dale did not intend to place value on one modality over another. *The shape of the cone is not related to retention, but rather to the degree of abstraction.*[1] However, he does contend that, as one's experiences move toward the bottom of the cone, more of the senses are engaged (such as hearing, seeing, touching, smelling, tasting).

Figure 5. Edgar Dale's Original Cone
In Dale's text, immediately before presenting the cone, he states:

"Much of what we found to be true of direct and indirect experience, and of concrete and abstract experience, can be summarized in a pictorial device which we call the 'Cone of Experience.' The cone is not offered as a perfect or mechanically flawless picture to be taken with absolute literalness in its simplified form. *It is merely a visual aid* [original italics] in explaining the interrelationships of the various types of audio-visual materials, as well as their individual 'positions' in the learning process…The cone device, then, is a visual metaphor of learning experiences, in which the various types of audio-visual materials are arranged in the order of increasing abstractness as one proceeds from direct experience…Exhibits are nearer to the pinnacle of the cone not because they are more difficult than field trips but only because they provide a more abstract experience. (An abstraction is not necessarily difficult. All words, whether used by little children or by mature adults, are abstractions.)"[2]



So where does the breakdown come from, and what is the real research behind it? An in-depth search of various citations produced countless dead ends. Sources quoted other sources that quoted still other sources, and the trail often led in circles. When authors were contacted directly, there was a tendency to insist, “I had the data once, but I don’t seem to have it in my files anymore.” One source said she “remembered it coming from some old Socony Mobil research,” but had no further information. In some cases, citations were incorrect and could not be followed up at all.

Further searching led to the work of a small group of researchers dedicated to debunking what they called this “bogus data.” Professor Frank Dwyer of Penn State, a noted expert on research in this area, criticized the data quite thoroughly in his book, *Strategies for Improving Visual Learning*.[3] Using Dwyer’s research as his starting point, Professor Michael Molenda from the University of Indiana also pursued the issue for several years. Molenda’s reader commentary yielded the most detailed historical accounting of the “research” yet. Molenda found evidence that the “data” were actually developed during World War II by Lieutenant Colonel Paul John Phillips, who prepared training materials for the Navy and the petroleum industry.

According to Molenda’s own research of historical records, before and after the war Phillips worked for the petroleum industry at the University of Texas (Austin) where he first prepared and distributed a handout with the “data.” Phillips’s work with the petroleum industry may explain how the “data” became connected with present-day Mobil Oil – also known in earlier years as Socony-Vacuum Oil Company and Socony Mobil Oil Company. While working as the head of Training Methods at the Ordinance School during the war, Phillips was responsible for the training of instructors and the development of training doctrine. The school’s historian, Pete Kindsvatter, verifies that Phillips served in this capacity, but has been unable to find any documentary evidence of the research on which the “data” disseminated was supposedly based. Nor does the University of Texas have any information about the studies on which Phillips’s “data” might be based.

Molenda concludes, based on his own searches and the research of others like him, that the “bogus data” are more representative of a “rounded-off generalization based on Phillips’s experience, and probably some test data, at the Ordnance School” than anything more substantial.[4 5 6] To date, and despite its widespread dissemination among scholars of all stripes, there is no more conclusive evidence of the data’s validity than that.

Why Do People Find the Cone of Experience so Compelling?

The complexity of today’s global society and the accelerating rate of change require a citizenry that continuously learns, computes, thinks, creates, and innovates. That translates into a critical need to become extremely efficient in the use of the time we spend learning, since we are being required to continuously learn throughout our lives.

One of the bottlenecks to efficient learning is our own physiology – the way our brains are wired severely limits our capacity to learn.[7 8 9 10] It is precisely this limitation that educators must overcome through informed design of learning environments, curricula, instruction, assessments, and resources. As they design lessons, create learning environments, and interact with students, they are seeking augmentations that accommodate for these human limitations. This is analogous to the design of machines (such as cars, tractors, elevators, robotic factories, can openers, stairs, etc.) used to accommodate for our severe physical strength and endurance limitations – only now we are augmenting intellectual capacity rather than physical capacity.

Educators are continuously redesigning learning experiences in order to increase and deepen learning for all students, as evidenced by the recent literature on differentiated learning.[11] Their efforts are much more likely to succeed when their work is informed by the latest research from the neurosciences (how the brain functions), the cognitive sciences (how people learn), and research on multimedia designs for learning.

The person(s) who added percentages to the cone of learning were looking for a silver bullet, a simplistic approach to a complex issue. A closer look now reveals that one size does not fit all learners. **As it turns out, doing is not always more efficient than seeing, and seeing is not always more effective than reading.** Informed educators understand that the optimum design depends on the content, context, and the learner. For example, the bogus percentages on the cone would suggest that engaging students in collaborative learning in general would result in higher levels of learning than would a lesson where a student listens to narration or reads text about the topic. The reality is that, for the novice student engaged in basic skill building such as learning chemical symbols, individual learning through reading or simple drill and practice might be the optimal learning design. Yet, for a different learning objective – for instance, understanding cause and effect of a specific chemical reaction – involving that same student in collaborative problem-solving with fellow students through a simulation might be the most effective learning approach.[12] [13]

Within those constraints, research is emerging that provides important guidelines for multimodal use of media with students. Optimizing learning for each student requires more fine-grained differentiation of instruction that takes into account – and leverages – each of the three areas mentioned earlier: *how the brain functions, how people learn, and multimedia design*.

Experienced teachers recognize that the design of lessons must adapt to the expertise and prior knowledge of the learner, the complexity of the content, and interests of the learner. Experienced researchers recognize that the use of technology and multimedia, resources, and lessons can vary in the level of interactivity, modality, sequencing, pacing, guidance, prompts, and alignment to student interest, all of which influence the efficiency in learning.[14]

The intent of this paper is to bring to light research findings on a critical aspect of Edgar Dale's Cone of Learning, the differential learning outcomes between single-mode (unimodal) and multiple modes (multimodal) of learning.

To provide the context for understanding that differential, this paper briefly summarizes key elements of emergent research in how the brain functions, how people learn, and prior research in multimodal learning. It then goes on to report meta-analytic findings on the multimedia principle – one of numerous considerations in multimodal learning. It concludes with implications for teachers in their design of lessons using media.

As background, definitions for learning, schema, and scaffolding are provided here.

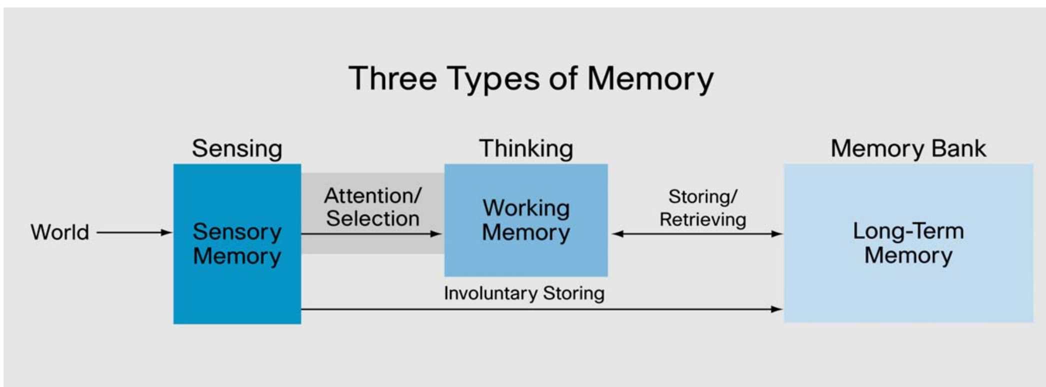
> **Learning** is defined to be "storage of automated schema in long-term memory."[15]
>
> **Schemas** are chunks of multiple individual units of memory that are linked into a system of understanding.[16]
>
> **Scaffolding** is the act of providing learners with assistance or support to perform a task beyond their own reach if pursued independently when "unassisted."[17]

1: How the Brain Functions – The Physiological Limitations to Learning

With time a limited commodity in today's society, people are tempted by technology to do more than one thing at a time (such as driving and talking on the phone, reading e-mails while participating in audio conferences, etc.). New scientific studies reveal the losses in efficiency in such multitasking. Researchers find that thinking processes happen serially, resulting in delays caused by switching from one task to another. The delays become more pronounced as the complexity of the task increases.[18] This explains why driver inattention and other human errors reportedly cause 40 percent of all traffic accidents.[19] One might ask why, with our incredibly sophisticated brain that uses 100 billion neurons to process information at rates of up to a thousand times a second, we are still incapable of doing two things at once? The answer is emerging from neuroscience labs around the world, where scientists are using fMRI and rapid sampling techniques to reveal the pattern of brain activity over time as people read, listen, talk, observe, think, multitask, and perform other mental tasks.[20]

Neuroscientists are reporting new discoveries that provide insights into long-held learning theories. For example, conjectures from decades ago on the existence of short-term and long-term memory[21] and cognitive overload[22] now have supporting evidence from the neurosciences.[23] Research indicates that the brain has three types of memory: sensory memory, working memory, and long-term memory.[24]

Figure 6. Memory Types



Based on Mayer (2003); Moreno & Mayer (2007); Marois (2005); and Miyake, et al (1999).

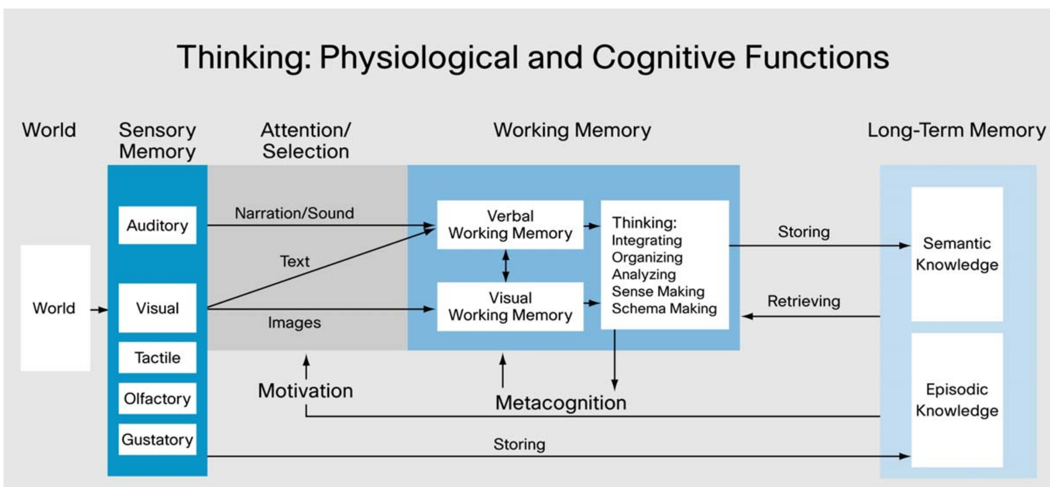
Three types of memory:

- **Working memory:** Working memory is where thinking gets done. While it is represented as a box in Figure 6, it is actually more brain function than location. The working memory is dual coded with a buffer for storage of verbal/text elements, and a second buffer for visual/spatial elements.[25] This represents one of the severe limitations of human thinking processes, for short-term memory is thought to be limited to approximately four objects that can be simultaneously stored in visual/spatial memory and approximately seven objects that can be simultaneously stored in verbal short-term memory. If those buffers are full and the person shifts attention, new elements may be introduced into working memory causing others to disappear from thought/consciousness. Within working memory, verbal/text memory and visual/spatial memory work together, without interference, to augment understanding. Overfilling either buffer can result in cognitive overload.[26] This includes buffers of visual/spatial memory traces and verbal (auditory and text) memory traces. Recent studies suggest that the brain is capable of multisensory convergence of neurons provided the sensory input is received within the same timeframe. Convergence in the creation of memory traces has positive effects on memory retrieval. It creates linked memories, so that the triggering of any aspect of the experience will bring to consciousness the entire memory, often with context.

- **Sensory memory:** Experiencing any aspect of the world through the human senses causes involuntary storage of sensory memory traces in long-term memory as episodic knowledge. These degrade relatively quickly. It is only when the person pays attention to elements of sensory memory that those experiences get introduced into working memory. Once an experience is in working memory, the person can then consciously hold it in memory and think about it in context.
- **Long-term memory:** The short-term memory acts in parallel with the long-term memory. Long-term memory in humans is unlimited estimated to store up to 10^9 to 10^{20} bits of information over a lifetime – equivalent to 50,000 times the text in the U.S. Library of Congress.[27] The brain has two types of long-term memory, episodic and semantic. Episodic is sourced directly from sensory input and is involuntary. Semantic memory stores memory traces from working memory, including ideas, thoughts, schema, and processes that result from the thinking accomplished in working memory. The processing in working memory automatically triggers storage in long-term memory.

Figure 7 maps the process of human thinking across the three memory buffers.

Figure 7. Schematic of the Thinking Processes



Based on Mayer (2003); Moreno & Mayer (2007); Marois (2005); and Miyake, et al (1999).

Consider the following example:

A learner is in a science lab, working in a team on the development of an architectural design related to geometry. The sights, sounds, tastes, and smells are involuntarily encoded in her sensory memory through her dual sensory channels (verbal/text and visual/spatial):

- **Verbal/text channel:** Side conversations, noise from other teams, bell systems, etc.
- **Visual/spatial channel:** Current architectural drawings on screen or paper, facial expressions, physical movements by others, etc.

  Note: Researchers believe that gustatory, olfactory, and tactile stimuli are logged through the visual channels, but there is less evidence as to the location of the storage buffers.

The involuntary memory traces are stored in long-term memory. As the student pays attention to various aspects of the sensory inputs, those inputs are also stored in short-term memory for a few seconds – lasting only as long as she causes the synapses to fire by thinking about the inputs (attention). As this student contemplates further about a particular side conversation related to traffic patterns within school in their architectural drawing and draws conclusions, the memory trace moves from short-term memory to long-term memory. As the student continues to contemplate the traffic pattern issues, she is also able to cue up memories from her own personal experiences (from long-term memory) that have enriched her thinking, and thus this new memory. Should she be distracted by something like an office announcement over the intercom, she may experience *attention blink* (AB) and lose sight of everything else around her due to the distraction in a specific or in multiple channels.[28]

During that experience, she might also have auditory overload that causes her to not register other discussions going on around her but that doesn't prevent her from continuing to register input involuntarily (which gets stored momentarily in long-term memory, but doesn't last long unless she pays attention to them, thus drawing them into short-term memory). Furthermore, as she consciously considers each sensory input or decides to work on a particular aspect of the architectural plans, her executive cognitive control function restricts her attention to serial consideration of ideas and concepts. Executive cognitive control is a phenomenon that slows down thinking and makes multitasking inefficient. While the student can simultaneously make a decision *and* continue to view the world around her *and* store memory traces in working/short-term memory (for these work in parallel); thinking, decision making, and cueing of long-term memories invoke and require the central cognitive processor, which only works serially. This is an important phenomenon for teachers to understand. Cognitive overload, dual processing, and the serial nature of the executive control explain the need for scaffolding of student learning.

2: How People Learn – The Cognitive Sciences

Research over the last two decades has revealed volumes on the subject of how people best learn. A 2001 publication from the National Academy of Sciences, *How People Learn*,[29] outlines important principles upon which schools should consider redesigning learning:

- **Student preconceptions of curriculum must be engaged in the learning process.** Students have preconceptions and prior experiences with many of the areas of study included in the academic standards. These are stored in long-term memory. Often some of those preconceptions turn out to be misconceptions. Student learning is greatly enhanced when each student's prior knowledge is made visible (that is, cued from long-term memory into working memory). It is at that point the student has the opportunity to correct misconceptions, build on prior knowledge, and create schemas of understanding around a topic. Learning is optimized when students can see where new concepts build on prior knowledge.
- **Expertise is developed through deep understanding.** Students learn more when the concepts are personally meaningful to them. In order to deeply understand a topic, learners not only need to know relevant facts, theories, and applications, they must also make sense of the topic through organization of those ideas into a framework (schema) of understanding. The development of schema requires that students learn topics in ways that are relevant and meaningful to them. This translates into a need for authentic learning in classrooms, (Note: Authentic learning is defined here to include three key concepts: depth of academic concept or deep learning, relevance to person(s) outside the classroom, and student use of the key ideas in a production.)

- **Learning is optimized when students develop "metacognitive" strategies.** To be metacognitive is to be constantly "thinking about one's own thinking," in search of optimizing and deepening learning. Students who are metacognitive are students who approach problems by automatically trying to predict outcomes, explaining ideas to themselves, noting and learning from failures, and activating prior knowledge. Given appropriate scaffolding by educators and other adults, all students can learn metacognitive strategies.

Despite recent advances, cognitive science is a relatively new field, and thus will undoubtedly continue to evolve as new research is conducted. New advances in functional magnetic resonance imaging (fMRI) have enabled cognitive sciences to look into the black box (that is, the brain) to investigate what have been up until recently, merely theories that fit patterns of behavior. That work will undoubtedly continue to evolve to inform educators.

The real challenge before educators today, is to establish learning environments, teaching practices, curricula, and resources that leverage what we now know about the limitations of human physiology and the capacity explained by the cognitive sciences to augment deep learning in students.

3: Multimedia Design – Visual and Verbal Learning

Recent neuroscience research is beginning to synergistically verify the previously speculative theories of multiple researchers in dual coding, cognitive overload, and multimedia learning.[30] While the field is still evolving, researchers have shown that significant increases in learning can be accomplished through the informed use of visual and verbal multimodal learning.

Much has been written about the principles of multimedia listed below. Most of the published research studies have been of short duration and were specifically designed for research analysis, but have demonstrated the veracity of these principles. However, emergent research on these principles, when applied in classrooms, has had mixed, albeit positive, results. Many of the researchers have commented that such mixed results may be due to the lack of specificity of the type of multimedia intervention (for example, specific combinations of modalities, formats within modalities, learner characteristics, scaffolding of learners, learner age, complexity and type of learning goals addressed, etc.)

A set of principles related to multimedia and modality are listed below. They are based on the work of Richard Mayer, Roxanne Moreno, and other prominent researchers.[31] [32] [33] [34]

1. **Multimedia Principle:** Retention is improved through words and pictures rather than through words alone.
2. **Spatial Contiguity Principle:** Students learn better when corresponding words and pictures are presented near each other rather than far from each other on the page or screen.
3. **Temporal Contiguity Principle:** Students learn better when corresponding words and pictures are presented simultaneously rather than successively.
4. **Coherence Principle:** Students learn better when extraneous words, pictures, and sounds are excluded rather than included.
5. **Modality Principle:** Students learn better from animation and narration than from animation and on-screen text.
6. **Redundancy Principle:** Students learn better when information is not represented in more than one modality – redundancy interferes with learning.

7a. I**ndividual Differences Principle:** Design effects are higher for low-knowledge learners than for high-knowledge learners.

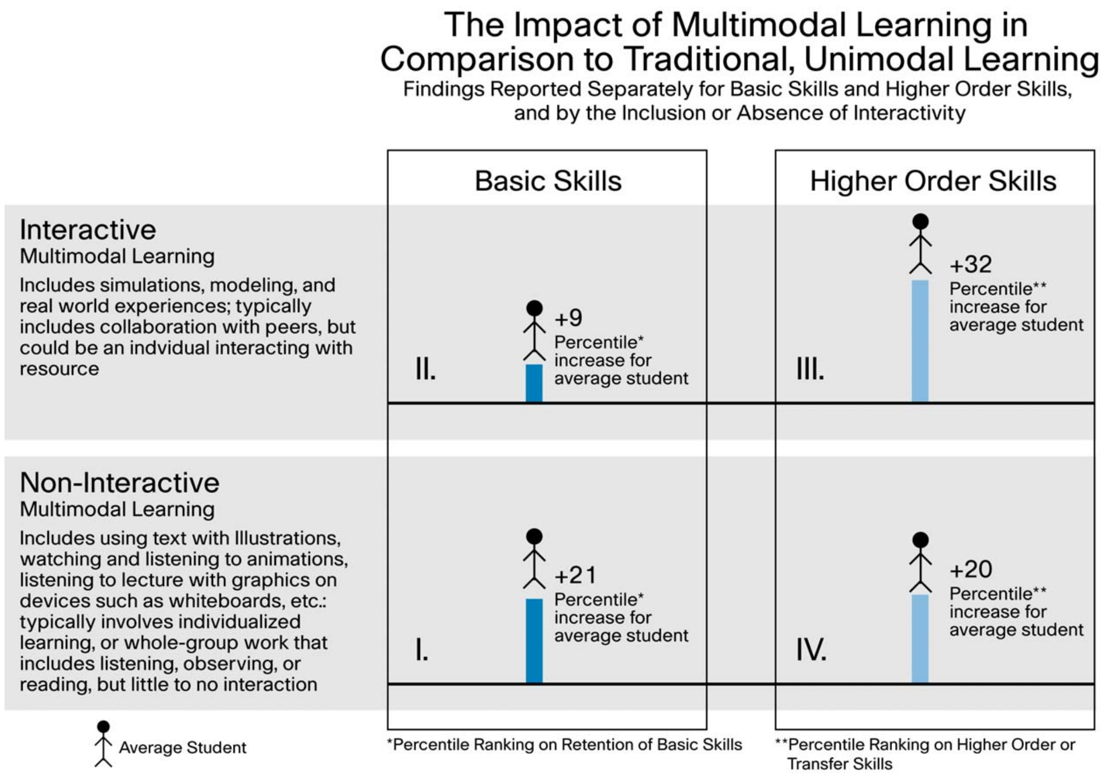
7b. **Individual Differences Principle:** Design effects are higher for high-spatial learners rather than for low-spatial learners.

8. **Direct Manipulation Principle:** As the complexity of the materials increase, the impact of direct manipulation of the learning materials (animation, pacing) on transfer also increases

New Web 2.0 technologies introduce some nuances to multimodal learning that warrant continued research. In practice educators are getting mixed, albeit positive trends in the use of multimedia to augment learning. **Students engaged in learning that incorporates multimodal designs, on average, outperform students who learn using traditional approaches with single modes.**

Figure 8 provides results from across multiple studies, separating effects related to basic and higher-order skills (see Appendix A for methodology and citations).

Figure 8. Impact of Multimodal Learning (Verbal and Visual)



The findings in Figure 8 are based on meta-analytic analysis and are summarized below:

- **Quadrants I and II:** The average student's scores on basic skills assessments increase by 21 percentiles when engaged in non-interactive, multimodal learning (includes using text with visuals, text with audio, watching and listening to animations or lectures that effectively use visuals, etc.) in comparison to traditional, single-mode learning. When that situation shifts from non-interactive to interactive, multimedia learning (such as engagement in simulations, modeling, and real-world experiences – most often in collaborative teams or groups), results are not quite as high, with average gains at 9 percentiles. While not statistically significant, these results are still positive.

- **Quadrants III and IV:** When the average student is engaged in higher-order thinking using multimedia in interactive situations, on average, their percentage ranking on higher-order or transfer skills increases by 32 percentile points over what that student would have accomplished with traditional learning. When the context shifts from interactive to non-interactive multimodal learning, the result is somewhat diminished, but is still significant at 20 percentile points over traditional means.

This analysis provides a clear rationale for using multimedia in learning. That said, the reader should be cautioned that the research in this field is evolving, with recent articles suggesting that efficacy, motivation, and volition of learners, as well as the type of learning task and the level of instructional scaffolding, can weigh heavily on the learning outcomes from the use of multimedia.[35 36 37]

Conclusion

The complexity of teaching and learning becomes increasingly apparent as the physiological, cognitive, social, and emotional aspects of learning become known. The percentages related to the cone of learning were a simplistic attempt to explain very complex phenomenon. The reality is that the most effective designs for learning adapt to include a variety of media, combinations of modalities, levels of interactivity, learner characteristics, and pedagogy based on a complex set of circumstances.

In general, multimodal learning has been shown to be more effective than traditional, unimodal learning. Adding visuals to verbal (text and/or auditory) learning can result in significant gains in basic and higher-order learning. The meta-analytic findings in this report provide insights into when interactivity augments multimodal learning of moderately to complex topics, and when it is advantageous for students to work individually when learning or building automaticity with basic skills.

Future Research

The opportunity for future original research and meta-analytic studies in this field is tremendous.

First, there continues to be opportunities to ask more specific research questions related to multimodal learning through high-tech media. Based on the meta-analytic findings in this report, another logical probe would be the differentiation between interactivity related to collaboration and that between a student and the software or Web resources.

The emphasis of the most multimedia studies to date has been on the impact on students' cognitive structures and processes only. Educators and researchers are now asking questions related to:

- **The social affordances that multimedia representations provide.** For example, Robert Kozma conducted research using multimedia representations in high school chemistry classes. His project simulated scientists' use of investigative laboratory activities to provide support discussions, studies, and argumentation that result in the construction of shared understandings of scientific phenomena.[38] Given the multiplicity of opportunity for social networking, collaborations, and student-student, student-instructor, and student-resource interactions, the complexities of the research need to become more specific and fine-grained.[39 40]
- **The scaffolding required to prepare students to effectively use multimedia, visual representations.** Many authors speculate that unless students have been trained to interpret visuals, the impact of multimedia will be minimal. For example, Roth and Bowen (2001) suggest that graph-related practices are skill sets that set scientists apart and are

highly contextualized. Thus students need to understand graphing specific to the phenomenon they are learning (such as the weather, demographics, traffic, chemistry, vectors, etc.)[41]

- **The learning designs necessary to minimize cognitive overload throughout the trajectory of the students' learning.** Cognitive load theory is concerned with techniques for managing working memory load in order to facilitate the changes in long-term memory associated with schema construction and automation. Sweller's theory of cognitive overload includes discussions of three types of loads in the working memory. Changes in long-term memory related to automaticity and schema construction are essential for managing the load on working memory.[42] The three types of memory load are:
    - **Intrinsic:** Memory that understands a concept or idea by establishing schemas (for example, interactivity between elements)
    - **Germane:** The degree of learner effort in construction of schemas, influenced by motivation and interest
    - **Extraneous:** Modality-specific neuron structures impacted by alignment between design elements and presentation

All three memory loads can now be measured through fMRI. The germane load represents the effort the person expends in constructing and storing schemas in long-term memory that represent learning.

As mentioned earlier, scaffolding is the provision of assistance to a learner in support of his/her performance that would otherwise be beyond his/her reach. Typically, the scaffolding is "faded," eventually enabling the learner to become fully accomplished in the task without the scaffolding. Roy Pea (2004) makes an important distinction between distributed intelligence, where scaffolding is integral to the task and won't be faded, versus scaffolded achievement, where fading occurs.[43] This is important given the increasing reliance on distributed intelligence among virtual teams versus individual intelligence; and the 24-hour reliance on distributed resources that is now commonplace in most work and many learning environments. For example, online resources such as search engines, browsers, dictionaries, and other resources are scaffolds for learning that probably will not be faded. This has interesting implications for assessments in schools that shift the emphasis to performance-based assessments of both individuals and teams.[44] [45] [46]

- **The importance of the attention and motivation of the learner.** Our propensity to pay continuous partial attention to multiple surroundings enables us to scan and rescan our environment. But to encode any of the observations into memory requires us to pay particular attention and to think specifically about that input. While we do involuntarily take in sensory input through our verbal and non-verbal channels, we can control what stays actively thought about in working memory, and thus, what gets stored in long-term memory (what gets learned). In addition to variations of impact on learning due to instructional design of learning experiences, there are variations based on learner expertise. The scaffolding of learning by reducing extraneous diversions, to focus the learner's attention on appropriate elements aligned to the topic, has proven effective.[47]
- **The importance of separating the media from the instructional approach.** One of the challenges in research on multimedia is the confound that occurs when the media and the pedagogy are not defined separately. A recent meta-analysis in which over 650 empirical studies compared media-enabled distance learning to conventional learning found pedagogy to be more strongly correlated to achievement than media.[48] [49]

The convergence of the cognitive sciences and neurosciences provides new insights into the field of multimodal learning through Web 2.0 tools. The combination will yield important guideposts in the research and development of e-learning using emergent, high-tech environments.

Appendix A: Methodology for Meta-Analytic Analysis

The intent of this review was to summarize across quantitative studies related to the effectiveness of multimodal learning in comparison to traditional learning. The multiple effect sizes used in this study were extracted from meta-analyses and experimental or quasi-experimental design studies published from 1997 to 2007.

Meta-Analysis Techniques

Meta-analysis statistical procedures provide a measure of the difference between two groups that is expressed in quantitative units that are comparable across studies. As Marzano, Pickering, and Pollock (2001) point out, "Being able to translate effect sizes into percentile gains provides for a dramatic interpretation of the possible benefits of a given instructional strategy."[50] Developed by Gene Glass in the mid-1970s, meta-analysis allows for comparisons to be made about the relative effectiveness of various strategies to increase student achievement.[51]

Included Studies

In the preliminary search for studies, the search terms: multimodal, modality effect, animation, multimedia, memory, retention, narration, and meta-analysis were used in various combinations. Many studies and reports were collected which proved to be unusable for this analysis. For final inclusion in the meta-analysis, more stringent criteria were applied. The initial search was for meta-analyses, with single studies included only if not included in the latest meta-analysis. To be included a meta-analysis or study had to have:

- Been published after 1997

  Note: If a meta-analysis was published after 1997 it was included even if the studies analyzed within the study were published before 1997.
- Addressed multimedia in education
- Used student achievement, retention, basic skills, higher-order skills, or transfer skills as dependent variable
- Been experimental or quasi-experimental comparing multimodal to single-modality learning
- Reported effect size (ES) or the statistics necessary to calculate
- Been for general education students, classes, or adults (not dealing exclusively with a special subpopulation)

For this analysis, a total of 14 studies or meta-analyses were identified through the application of the above criteria. Within the 14 articles, 23 independent studies or meta-analyses were identified. Each of the independent studies was then classified based on the type of intervention (multimodal interactive or multimodal non-interactive) and by the type of assessment used to determine the effect size (assessment of retention/basic skills or assessment of higher-order/transfer skills). That enabled the researchers to classify the studies into four categories: 1) Basic skills, non-interactive; 2) Basic skills, interactive, 3) Higher-order skills, interactive, and 4) Higher order skills, non-interactive. (See tables A1–A4.) Overall, the studies contained a combined sample of nearly 6,000 students. The average number of students in each independent study or meta-analysis was approximately 260. Seventy percent of the studies were published after 2003, while 30 percent were published before.

Methods for Calculating Effect Size

Effect sizes were calculated using sample and study-weighted procedures. Sample weighting means greater weight is given to effect sizes associated with larger samples based on the assumption that larger samples are better able to approximate actual effects of the target population. In other words, the weighted effect size equals the sum of the products of the study effect sizes (*d*) and their associated sample sizes (n), divided by the sum of the n's.

The studies used in the meta-analysis varied by the number of comparisons they reported. These multiple results from the same study can be problematic for meta-analysis because the separate estimates in the same study are not completely independent – they share historical and situational influences, and some of them even share influences contributed by having been collected from the same people.[52] To give all studies the same unit weight in the analysis, only one effect size was used from each independent study. In some cases this represented the average of appropriate effect sizes; in other cases it was simply the effect size appropriate to the analysis criteria.

The following four tables include the studies analyzed based on the categorization described above.

Table 1. Studies for Non-Interactive, Basic Skills (Quadrant 1)

Non-Interactive Multimodal Basic Skills	Intervention	Assessment	N	ES
Moreno & Valdez (2005)	Non-interactive (words vs. words/pictures + pictures vs. words/pictures)	Retention	35	1.30
Chan & Black (2006)	Text vs. text + static visuals	Recall (visual and verbal)	189	0.47
Kalyuga et al (2004)	Auditory only vs. visual + auditory	Retention test	21	–0.82
Kim & Olaciregui (2007)	Folder-based vs. concept map access to digital resources	Retention	51	0.81
Kalyuga et al (1999)	Auditory vs. visual vs. auditory + visual	Multiple choice – content	16	0.74
Dubois & Vial (2000)	Text vs. integration of text/image/sound	Vocabulary test	45	0.61

Table 2. Studies for Interactive, Basic Skills (Quadrant II)

Interactive Multimodal Basic Skills	Intervention	Assessment	N	ES
Moreno & Valdez (2005)	Interactive (words vs. words/pictures + pictures vs. words/pictures)	Retention	40	1.77
Lee (1999)	Meta-analysis: simulations	Achievement	563	0.35
Rosen & Salomon (2007)	Meta-analysis: constructivist, technology-rich	Retention	2168	0.11
Kalyuga et al (2004)	Exp. #1 – Text vs. visual + audio	Multiple choice – content	34	–0.64
Chan & Black (2006)	Text vs. text + manipulations	Recall (visual and verbal)	189	1.04

Table 3. Studies for Interactive, Higher-Order/Transfer Skills (Quadrant III)

Interactive Multimodal	Intervention	Assessment	N	ES
Chan & Black (2006)	Text vs. text + manipulations	Transfer (averaged)	189	0.84
Moreno & Valdez (2005) NOTE: Averaged comparison of text and direct manipulation to text	Interactive (words vs. words/pictures + pictures vs. words/pictures)	Transfer	35	1.06
Grimshaw et al. (2006)	CD text vs. CD text + narration	Comprehension test	52	0.68
Kim & Olaciregui (2007)	Folder-based vs. concept map access to digital resources	Comprehension tests	51	1.42
Mayer (2003)	Meta (5)	Comprehension tests	204	1.67
Rosen & Salomon (2007)	Meta-analysis: constructivist, technology-rich	Transfer	1837	0.90
De Westelinck (2004)	Compared text to text + text representation	Retention/transfer	63	–1.12
Atkinson (2002)	Control vs. text + agent or voice + agent	Far transfer	30	0.82

Table 4. Studies for Non-Interactive, Higher-Order/Transfer Skills (Quadrant IV)

Non-Interactive Multimodal Higher Order/Transfer Skills	Intervention	Assessment	N	ES
Chan & Black (2006)	Text vs. text + static visuals	Transfer (averaged 3 transfer effect sizes)	189	0.40
McKay (1999)	Text only vs. text + graphics	Pre-/post-test	41	0.53
Moreno & Valdez (2005)	Non-interactive (words vs. words/pictures + pictures vs. words/pictures)	Transfer test	40	0.84
Tindall-Ford, Chandler, and Sweller (1997)	Text vs. text + graphics	Transfer test	30	0.99

Note: Cohen (1969) proposed that an effect size of 0.2 could be regarded as “small,” that an effect size of 0.5 could be regarded as “medium,” and that an effect size of 0.8 could be regarded as “large.” Cohen’s proposals have become widely accepted as rules of thumb in educational and social research. However, they were put forward without any consideration of the effect sizes that it was reasonable to expect in real research studies.

Appendix B: References for Studies Included in Meta-Analysic Analysis

Atkinson, R.K. (2002). Optimizing Learning From Examples Using Animated Pedagogical Agents. Journal of Educational Psychology. Vol. 94, No. 2, 416-427.

Chan, M. S., & Black, J. B. (2006). Learning Newtonian mechanics with an animation game: The role of presentation format on mental model acquisition. Paper presented at the American Education Research Association Annual. San Francisco, CA.

De Westelinck, K, Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in social sciences: limitations of external graphical representations. Computers in Human Behavior. 21 (pp. 555-573).

Dubois, M. & Vial, I. (2000). Multimedia design: the effects of relating multimodal information. Journal of Computer Assisted Learning. Vol. 16, 157-165.

Grimshaw, S., Dungworth, N., McKnight, C., & Morris, A. (2006). Electronic books: children's reading and comprehension. British Journal of Educational Technology, 38(4), 583-599.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psycholog*y, 13(4), 351-371.

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When Redundant On-Screen Text in Multimedia Technical Instruction Can Interfere with Learning. Human Factors, 46(3), 567-582.

Kim, P., & Olaciregui, C. (2007). The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere. British Journal of Educational Technology.

Lee, J. (1999). Effectiveness of Computer-Based Instructional Simulation: A Meta Analysis. *International Journal of Instructional Media*, 26(1), 71-72.

Mayer, R. E. (2003). Elements of a Science of e-Learning. *Journal of Educational Computing Research*, 29(3), 297-313. McKay, E. (1999). An Investigation of Text-based Instructional Materials Enhanced with Graphics. Educational Psychology, 19(3), 323-335.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. Educational Technology Research and Development, 53(3), 35-45.

Rosen, Y. & Salomon, G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: a meta-analysis. Journal of Educational Computing Research. Vol. 36(1) 1-14.

Rosen, Y. (2007). Personal correspondence via email yielded sample sizes for studies used in 2007 meta-analysis.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. Journal of Experimental Psychology: Applied, 3(4), 257-287.

[1] Dale, E. (1946, 1954, 1969). Audio-visual methods in teaching. New York: Dryden.

[2] Dale, E. (1954). Audio-visual methods in teaching. New York: Dryden. p. 42.

[3] Dwyer, F. M. (1978). *Strategies for Improving Visual Learning: A Handbook for the Effective Selection, Design, and Use of Visualized Materials*: Learning Services.

[4] Molenda, M. Cone of Experience.

[5] Technology, E., & Learning, V. Reader Comment: On the Origins of the "Retention Chart" An addendum to Subramony.

[6] Thalheimer, W. Website: http://www.work-learning.com/chigraph.htm, accessed 9/8/07.

[7] Dux, P. E., Ivanoff, J. G., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. Neuron, 52(6), 1109-1120.

[8] Fougnie, D., & Marois, R. (2006). Evidence From Attentive Tracking and Visual Working Memory Paradigms. *Psychological Science*, 17(6), 526-534.

[9] Miyake, A., & Shah, P. (Eds.) (1999). Models of working memory: Mechanisms of active maintenance and executive control. Cambridge University Press: New York.

[10] Baddeley, A. D. (1997). *Human Memory: Theory and Practice*: Psychology Press.

[11] Tomlinson, C.A., & McTighe, J. (2006). Integrating differentiated instruction and understanding by design. Alexandria, VA: ASCD (200 pp.).

[12] Bransford, J., Brown, A., & Cocking, R. (2000). How people learn. National Academy Press: Washington, DC.

[13] Pea, R. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. The Journal of the Learning Sciences, 13(3), 423-451. Lawrence Erlbaum Associates, Inc.

[14] Clark, R., Nguyen, F., and Sweller, J. (2006). Efficiency in Learning: Evidence-based Guidelines to Manage Cognitive Load. Pfeiffer.

[15] Sweller, J. (2002). Visualisation and instructional design. Paper presented at the International Workshop on Dynamic Visualisations and Learning, Tubingen.

[16] Bransford, J., Brown, A., & Cocking, R. (2000). How people learn. Brain, Mind, Experience, and School. Expanded version. National Academy Press: Washington, DC. p.33.

[17] Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17,.p.90, as cited by Pea, R. D. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *Scaffolding: A Special Issue of the Journal of the Learning Sciences*.

[18] Rubinstein, J.S., Meyer, D.E., & Evans, J.E. (2001). Executive Control of Cognitive Processes in Task Switching. University of Michigan, Ann Arbor, Mich., Journal of Experimental Psychology - Human Perception and Performance, Vol 27. No.4.

[19] Marois, R. (2005a). Capacity limits of information processing in the human brain. Phi Kappa Phi Forum, 85, 30-33.

[20] Marois, R. (2005b). Two-timing attention. Nature Neuroscience, 8, 1285-1286.

[21] Baddeley, A.D. (1992). Working Memory. Science; 255-256.

[22] Sweller, J. (1988). Cognitive load during problem solving, Cognitive Science 12: 257–285.

[23] Marois 2005a.

[24] Marois, R. & Ivanoff, J. (2005). Capacity limits of information processing in the brain. Trends in Cognitive Sciences, 9(6), 296-305.

[25] Baddeley, A. D. (1997). *Human Memory: Theory and Practice:* Psychology Press.

[26] Sweller, J. (1999). Instructional Design in Technical Areas. Melbourne, Australia: ACER Press.

[27] Marois 2005a.

[28] Marois & Ivanoff, 2005.

[29] Bransford et al (2000).

[30] Fougnie, D., & Marois, R., 2006.

[31] Mayer, R. (2001). Multi-Media Learning. Cambridge University Press.

[32] Mayer R.E. & Moreno R. (2003) Nine ways to reduce cognitive load in multimedia learning. In Web-Based Learning: What Do We Know? Where Do We Go? (eds R. Bruning, C.A. Horn & L.M. PytlikZillig), pp. 23–44. Information Age Publishing, Greenwich, CT.

[33] Chan, M. S., & Black, J. B. (2006). Learning Newtonian mechanics with an animation game: The role of presentation format on mental model acquisition. Paper presented at the American Education Research Association Annual. San Francisco, CA.

[34] Ginns, P. (2005). Meta-analysis of the modality effect. Learning and Instruction. Vol. 15, pp. 313-331. Institute for Teaching and Learning, University of Sydney, Australia.

[35] Clark, R.E. & Feldon, D.F. (2005). Five Common but Questionable Principles of Multimedia Learning. In Mayer, R. (Ed.) Cambridge Handbook of Multimedia Learning. Cambridge: Cambridge, University Press.

[36] De Westelinck, K, Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in social sciences: limitations of external graphical representations. Computers in Human Behavior. 21 (pp. 555-573).

[37] Deimann, M. & Keller, J. (2006). Volitional Aspects of Multimedia Learning. Journal of Educational Multimedia and Hypermedia. 15(2)137-158.

[38] Kozma, R. (2003). The Material Features of Multiple Representations and their Cognitive and Social Affordances for Science Understanding. Learning and Instruction, 13(2), 205-226.

[39] Michalchik, V., Rosenquist, A., Kozma, R., Kreikemeier, P., & Schank, P. (in press). Representational Resources for Constructing Shared Understandings in the High School Chemistry Classroom. To appear in J. Gilbert, M. Nakhleh, & M. Reiner (eds.), Visualization: Theory and practice in science education. New York: Springer.

[40] Roth, W.-M., & Bowen, G. M. (1999). Complexities of graphical representations during ecology lectures: an analysis rooted in semiotics and hermeneutic phenomenology. Learning and Instruction, 9, 238–255.

[41] Roth & Bowen (2001).

[42] Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, 38(1), 1-4.

[43] Pea 2004.

[44] Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing Split-attention and Redundancy in Multimedia Instruction. Applied Cognitive Psychology. Vol. 13, 351-371.

[45] Brunye, T., Taylor, H., Rapp, D., & Spiro, A. (2006). Learning Procedures: the Role of Working Memory in Multimedia Learning Experiences. Applied Cognitive Psychology (20) 917-940. Wiley InterScience.

[46] Pea 2004

[47] Paas, Renkl, & Sweller, 2004

[48] Clark, R.E. & Feldon, D.F., 2004

[49] Bernard, R.M., Abrami, P.C., Lou, Y., Borokhovs.ki, E., Wade, A., Wozney, L., Wallet, P.A., Fiset, M., & Huang, B. (2004). How Does Distance Education Compare to Classroom Instruction? A Meta-analysis of the Empirical Literature. Review of Educational Research, 74(3), 379-439.

[50] Marzano, R.J., Pickering, D.J., and Pollock, J.E. (2001). Classroom Instruction That Works: Research-Based Strategies for Increasing Student Achievement. Alexandria, VA: Association for Supervision and Curriculum Development.

[51] Glass, G. V. (1981). Meta-analysis in social research: Sage Publications Beverly Hills.

[52] Cooper, H. (1998). Synthesizing Research: a guide for literature reviews: Sage Publications Inc.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-455373-00 03/08