Measuring the Effectiveness of Educational Technology: what are we Attempting to Measure?

Jodie Jenkinson
University of Toronto at Mississauga, Canada
j.jenkinson@utoronto.ca

Abstract: In many academic areas, students' success depends upon their ability to envision and manipulate complex multidimensional information spaces. Fields in which students struggle with mastering these types of representations include (but are by no means limited to) mathematics, science, medicine, and engineering. There has been some educational research examining the impact of incorporating multiple media modalities into curriculum specific to these disciplines. For example, both Richard Mayer (multimedia learning) and John Sweller (cognitive load) have contributed greatly to establishing theories describing the basic mechanisms of learning in a multimedia environment. However when we attempt to apply these theories to the evaluation of e-learning in a more dynamic "real world" context the information processing model that forms the basis of this research fails to capture the complex interactions that occur between the learner and the knowledge object. It is not surprising that studies examining the effectiveness of e-learning technology, particularly in the area of basic science, have reported mixed results. In part this may be due to the quality of the stimuli being assessed. This may also be explained by the context in which interactivity is being utilized and the model that is used to evaluate its effectiveness. Educational researchers have begun to identify a need for more fine-grained research studies that capture the subtleties of learners' interactions with dynamic and interactive learning objects. In undergraduate medical and life science education, interactive technology has been integrated into the curriculum at many levels. This paper reviews experimental studies drawn from personal experience where an attempt has been made to measure the efficacy of educational technology. In examining the shortcomings of these more traditional experiments, we can then apply this understanding to characterizing a more flexible approach to evaluation and its potential in measuring the effectiveness of educational technology. Understanding the nature of technology-mediated learning interactions and the way in which they foster depth of understanding is a great challenge for both educational researchers and developers of e-learning technologies. By adopting an evaluative framework that takes a more flexible approach to measuring the emergent nature of understanding, we can examine the capacity of educational technology to support more complex understanding of curricular subject matter.

Keywords: science, e-learning, educational technology, evaluation, multimedia

1. Introduction

With the emergence of innovative electronic teaching and learning tools, technology has radically altered the surface of the educational landscape. From simply mining the Web for information, to engaging in simulated experiences, we increasingly situate educational technology as the driving force in learning. As we continue to integrate technology into teaching practice, we struggle with understanding the true value of these various media modalities in learning. Educational technology is a somewhat generic term that describes both the study and process by which technology may be used to advance learning. Scardamalia (2006) describes three distinct areas of technology that have potential implications for contributing to depth of understanding. These include: 1) Computer-assisted instruction (CAI); 2) Simulations, games, and laboratory instruments; and 3) Technology to support discourse. In particular, the use of CAI to complement traditional teaching has become a common feature of post-secondary education. However, the degree to which current uses of technology-assisted instruction contribute to deep understanding, has oftentimes proved difficult to measure.

There has been some educational research examining the impact of incorporating multiple media modalities into curriculum. In particular, Richard Mayer (1991, 1998, 1999, 2003) has contributed to establishing a cognitive theory of multimedia learning, which builds upon assumptions of how individuals learn. Firstly, Mayer asserts Paivio's theory (1983) of dual coding, one that postulates humans possess separate channels for processing visual and auditory information. Secondly, Mayer notes that humans have a limited capacity for the amount of information that each channel can process at one time. Lastly, he asserts that individuals learn by active engagement with cognitive processes, such as the selection, organization and integration of information (sensory memory, working memory, and long-term memory). Mayer's cognitive theory of multimedia learning addresses both the strengths and limitations of human perception and cognition is closely linked to John Sweller's (1988) cognitive load theory. Sweller describes the limitations of working memory and devises instructional techniques to facilitate the acquisition of knowledge in long-term memory.

Reference this paper as:
Jenkinson, J. "Measuring the Effectiveness of Educational Technology: What are we Attempting to Measure?" *Electronic Journal of e-Learning Volume 7 Issue 3 2009, (pp273 - 280), available online at www.ejel.org*

Cognitive load theory provides a framework for instructional design by distinguishing between 3 types of cognitive load (intrinsic, extraneous, and germane) and their association with learning. Intrinsic cognitive load has been described by Sweller and Chandler (1994) as arising from the interaction between the learning material and the expertise of the learner. Extraneous load is the cognitive load that extends beyond the intrinsic, and germane cognitive load is the load devoted to processes related to the construction and automation of schemas (Sweller, Van Merriënboer, and Paas, 1998). While intrinsic load is fixed, extraneous load and germane load may be directly impacted upon by instructional design (Paas, Ayres, and Pachman, 2008). Hence, experiments measuring cognitive load are often used to evaluate the success (or failure) of technology in reaching its audience. Both Mayer and Sweller's research have contributed greatly to establishing theories describing the basic mechanisms of learning in a multimedia environment. However, when we attempt to apply these theories to the evaluation of multimedia in a more dynamic "real world" context, the information processing model that forms the basis of this research, and the traditional methods of measurement, both fail to capture the complex interactions that occur between the learner and the subject matter.

2. Evaluating interactive media

2.1 How are we measuring the effects of interactive media?

In a research paradigm that attempts to measure change, the gold standard is the experimental design model. Accordingly, the evaluation of educational technology involves the randomization of students into one of 2 treatment groups: control and experimental. Measurement in the form of a pre-test establishes a baseline for evaluating the efficacy of the tool. Students are exposed to the intervention and this is followed by a post-test. Any significant change between pre and post is reported and attributed to the intervention. Certainly this is a model with which we are all familiar and within which many of us have conducted research with varying degrees of success. Those who argue in favour of a quantitative approach to evaluating educational technology do so on the grounds that it produces reliable and ecologically valid results that are readily generalisable. Proponents of a qualitative approach to evaluating educational technology, would argue that qualitative methodology is a more sensitive form of measurement; one that generates richer, more meaningful results. This is certainly not a new argument. As Oliver (2000) notes, the 'paradigm debate' is perhaps one of the longest running discussions within the evaluation community. It is true that there are advantages and disadvantages associated with either research paradigm. Neither has been successful in arguing its merits over the other. Robinson and Schraw (2008) identify the need for "quality research" in e-learning, citing a number of scientifically-based research studies that make unsupported claims about the benefits of e-learning. Many of these claims arise from flawed experimental design, erroneous or non-statistically significant effect-size comparisons, or purely subjective measures. Similarly Reeves (2007) is critical of the abundance of "one-off" quasi-experimental studies that are not linked to any particular research agenda. However the problem of evaluation is not limited to the mismeasure (either qualitative or quantitative) of e-learning. At the crux of this debate is a question (one that is too often overlooked) of precisely what it is that we are attempting to measure when we evaluate educational technology?

2.2 What are we measuring?

Typically, studies measuring the impact of educational technology are examining either the efficacy of the tool in teaching students, or the end-user's interaction with the system. Whereas efficacy is generally measured in terms of knowledge gain, usability studies are concerned primarily with the functionality of the device, regardless of whether or not learning objectives are being met. There is much research spanning a number of disciplines that examines successful approaches to measuring usability. Our attention here will be on the measurement of knowledge and understanding.

When initially we set out to evaluate the impact of technology upon learning, more often than not we are attempting to compare the benefits of a technological innovation with traditional pedagogy. Success of the technology is measured in terms of student performance, as demonstrated by tests assessing factual recall and knowledge of basic concepts. While it seems reasonable to assume that this is an accurate indicator of success, it often fails to tell us a great deal about the student's interaction with the learning tool. In other words, while it may tell us *what* new knowledge is being learned, it tells us nothing of *how* new knowledge has developed. Furthermore, traditional assessments frequently fail to detect a significant difference between treatments. It may be that in such cases there truly is no difference, and that as Clark (1994) once suggested, the media has little to do with learning outcomes. Or, it can be argued that media does play a role in learning and that we're just asking the wrong questions.

This problem of assessment is not limited to studies comparing traditional teaching methods with technology-enhanced teaching. Technology-to-technology comparisons are similarly difficult to assess and are plagued by a history of *no significant differences* (Reeves, 2007). Multimedia environments tend to be highly complex, containing a number of interacting variables. This poses a significant challenge when one attempts to assess the impact of educational technology upon learning. The standard approach to managing this complexity is to strictly control the manipulations to the variables being compared. For example, in two related studies (Jenkinson et al., 2007; Stewart et al., 2008) examining the effects of varied media modalities upon students understanding of dynamic processes, 154 first-year biology students were exposed to two e-learning modules, one of which contained animated graphics and the other containing static graphics. In every other respect the programs were identical. The purpose of this study was to identify whether animation was more effective than static graphics at teaching neurotransmitter release. A subsequent study (n=65) compared the efficacy of animated media with interactive media in teaching the same dynamic processes (both are illustrated in Figure 1). Both studies followed a structure that included pre-test, followed by time-limited exposure to one of 2 treatments, and then post-test. Neither experiment detected a significant difference between treatments (that isn't to say that we didn't see differences in the data; those differences were just not measurably significant). Interestingly, while the quantitative data failed to yield significant results, qualitative data (feedback forms and focus group evaluation) showed remarkable perceived differences in students' perception of the effectiveness of the media with which they engaged. Unfortunately, our research methods were not tightly integrated enough to explain this discrepancy. Similar studies examining factors such as timeline pacing in animated media (Visscher et al., 2009), and the placement of embedded self-examination within animated media have correspondingly demonstrated no difference in treatment effect, but measurable differences in user perceptions (Lui et al., 2006). While the results of these studies would suggest that media modality does not influence learning, there is evidence in the literature suggesting, to the contrary, that it does have a positive impact upon learning (reviewed in Anglin et al., 2004; Hidrio and Jamet ,2008; Ainsworth, 2008; Tasker and Dalton, 2008).

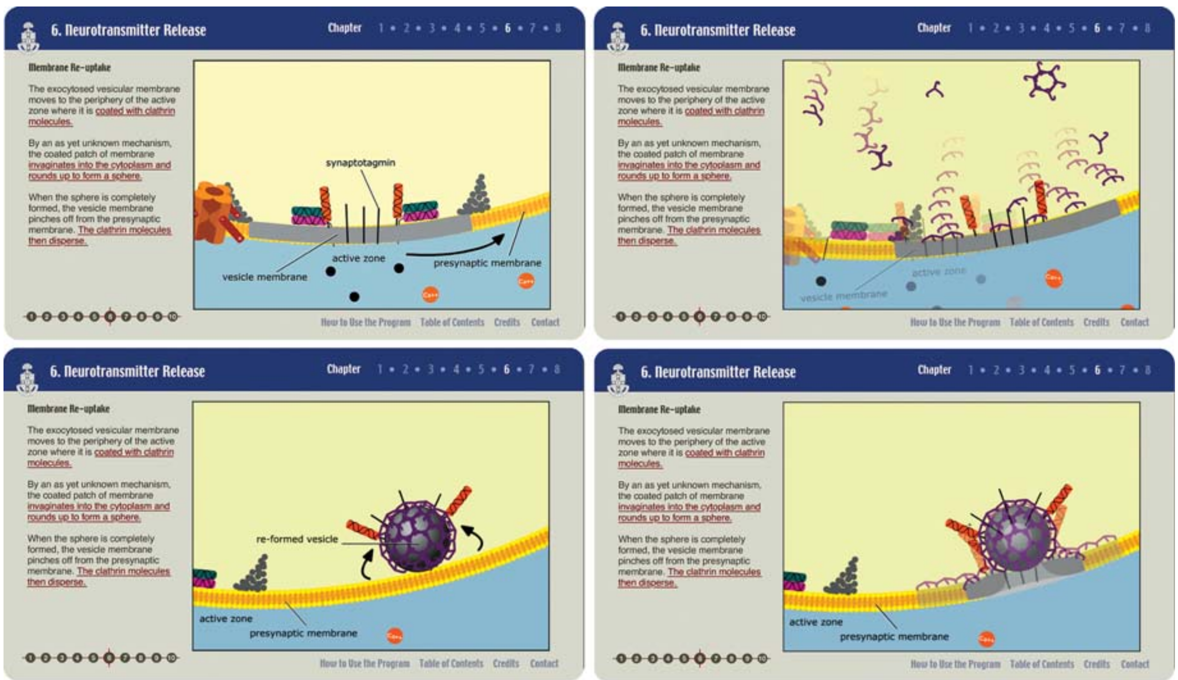

Figure 1: Depicting dynamic concepts with static (left) and animated (right) media (Jenkinson, Stewart & Cameron, 2007)

In another example, a study measuring the efficacy of a three-dimensional model in teaching functional human anatomy to undergraduate students (n=80), comparisons are made between static cardinal anatomical views and a fully rotational model of the pterygopalatine fossa (Kryski, 2008; illustrated in Figure 2).

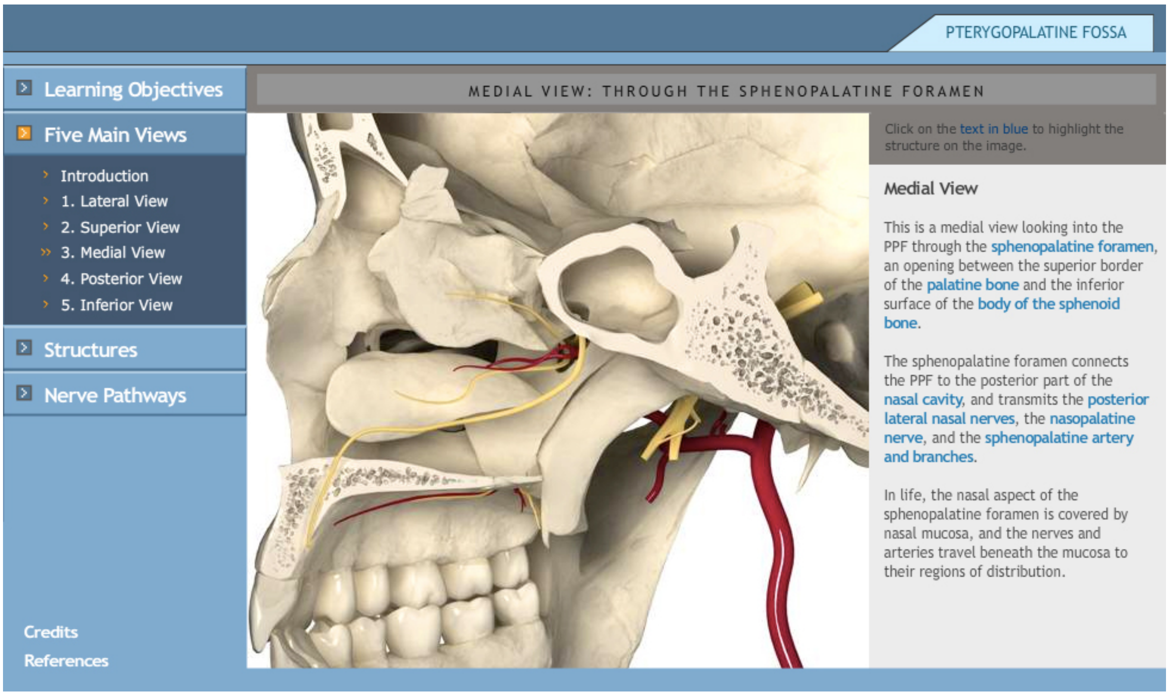

Figure 2: Screen-capture of web-based three-dimensional model of the pterygopalatine fossa (Kryski, 2008)

While this particular study failed to demonstrate a significant difference between treatments, similar studies examining the effectiveness of interactive 3D models in the study of anatomy have reported significant, but mixed results (Garg et al., 2002; Luursema et al., 2006; Nicholson et al., 2006). As with the findings associated with animated two-dimensional media, the findings reported by these studies are not surprising. In part this may be due to the quality of the stimuli being assessed. This may also be explained by the context in which interactivity is being utilized and the model that is used to evaluate its effectiveness. For example, in a study of a computer-based 3D model of the carpal bones of the hand, Garg et al. (2002) concluded that computer-based, manipulable, three-dimensional models are no more effective than static views in teaching complex spatial anatomy (in some cases they may even detract from learning). This is attributed to students' tendency to gather important spatial information from several key views only. Thus, time spent studying non-essential oblique views effectively reduced students' learning time. As the authors note, however, given the arrangement of the carpal bones (they naturally lie in two planes and lend themselves readily to two-dimensional representation), the object of study might not have been appropriate. In other words, the viewer gains very little new information about the carpal bones from side or oblique views. In this particular case it would appear as though the selected media modality (3-dimensional rotational model) is poorly matched with the learning objective (understanding the 2-dimensional planar arrangement of structures). As well, it may be that the data collection method (experimental design incorporating pre/post multiple choice tests) did not capture adequately *how* the students were learning from the interactive model.

3. A more flexible approach to assessment

In our efforts to measure the efficacy of educational technology it would appear as though we are at times sacrificing an opportunity to explore understanding in a more meaningful way, in favour of more replicable, generalisable results. To reiterate a concern expressed in the previous section of this paper, while this model of evaluation may tell us *what* new knowledge is learned by students, it fails to describe the transformative process by which new knowledge develops, and the factors involved in supporting and sustaining this change. If we are to create truly rich interactive experiences, we need to attend more closely to the contents of the 'black box' that is understanding. It is important that we distinguish between knowledge and understanding, and recognize that while knowledge may be more readily captured with traditional methods of evaluation, understanding, given its emergent nature, is more elusive.

3.1 Asking the right questions

Researchers in education (Ploetzner and Lowe, 2004; Ainsworth, 2008) have begun to identify a need for more fine-grained research studies that capture the subtleties of learners' interactions with dynamic and interactive tools. Shaaron Ainsworth (2008) has remarked that while some "first generation" experiments have been successful in producing robust and replicable results they fail to answer four important questions: 1) Who benefits from learning with (specific forms of) multimedia?; 2) How do people learn with multimedia?; 3) How does learning with multimedia change over time?; and 4) How does the wider context influence learning with multimedia? In order to answer these questions and capture the process by which learners interact with multimedia Ainsworth suggests that we should explore different, perhaps more flexible forms of evaluation design.

Robson's (2002) discussion of 'real world' research is an informative introduction to flexible research design in applied settings. Temporal and contextual factors, and questions such as *who* learns, and *how* we learn may be addressed using a mixed methods approach that combines quantitative research with qualitative data collection techniques such as ethnography, case study, phenomenology, cognitive task analysis, or microgenetic evaluation. For example, in a study examining how students interact with user-controllable animations while engaging in learning tasks, Lowe (2008) describes the effective use of combined qualitative and quantitative data sets that tightly integrate concurrent and retrospective verbalisations. "Think-aloud" protocols have proven very effective in eliciting user response to interactive systems, and in identifying important aspects of the novice-expertise continuum. Educational psychologists, perhaps most notably Siegler (see Siegler and Crowley, 1991) have used microgenetic methods for a number of years in examining the mechanisms that produce change. Microgenetic data sampling involves making a high rate of observations relative to the rate of change. It is an effective means of measuring change while it is occurring rather than examining pre- and post-change effects. More recent approaches to evaluation have combined these techniques with measurements of physiological changes (such as brain or eye activity). Eye tracking, for example, is used to index eye movements that occur when an individual is exposed to different visual environments (often while the user is completing a task). It is frequently used in combination with concurrent or retrospective verbal protocols. Eye tracking is well suited to providing a detailed account of attentional processes elicited by various multimedia representations, and possibly helping to explain how known learning effects (such as split-attention, or goal specificity) occur (Van Gog and Scheiter, 2009). Eye tracking has also been used effectively to compare novice and expert interactions with multimedia (Jarodzka et al., 2009; Van Gog, Paas, and Van Merriënboer, 2005). The various data collection methods that have been described here are suggested as possible means of accessing the proverbial black box. They are by no means a panacea for understanding the complex interactions of learners with educational technology. The shear richness and dimensionality of that experience is what makes it so difficult to assess.

The question of how to assess would appear to be two-fold: 1) How can we measure the impact of technology-mediated instruction in a way that is sensitive enough to detect the its role in fostering understanding?; and 2) How can we do this is in a way that is both reliable, valid, and to some extent transferable? The point of this discussion is not to suggest that we abandon quantitative research methods but rather, that we thoughtfully integrate multiple methodologies and data sources in evaluating educational technology. Complementary exploratory, and experimental studies are necessary to characterize the learning that occurs as a result of complex interaction with educational technology.

3.2 Characterising flexible design

In proposing that we take a more integrative or flexible approach to evaluating educational technology, the suggestion here is that we adopt a research paradigm that sits somewhere between traditional randomized trials and qualitative research, affording reiteration and revision of measures as necessary in order to better understand the learning situation. Vicente (1999, 2004) has suggested that, in order to capture the dynamics of the human-technology relationship, we need to think of that relationship as a system, to be examined holistically. He further points out that this relationship is not a physical property of the system, but rather an emergent property, "a gestalt, which only comes into existence when the parts it comprises are brought together and configured in a particular way" (2004, p. 46). Capturing the emergent nature of that relationship, in order that we might answer these questions, demands a multi-faceted learner-centred approach to evaluation; one involving a range of methods and measures.

One such methodology, that is gaining popularity, is design-based research (Brown, 1992; Collins, 1992). Learning scientists engaged in design experiments would describe this research model as an extended and refined process of investigation based upon principles derived from prior research (Collins, Joseph, and Bielaczyc, 2004; Confrey, 2006). Whereas the goal of structured laboratory studies is to control for single variables, design-based research attempts to describe the system as a set of interdependent elements, recognizing that the system in which learning naturally occurs is, for lack of a better term, messy. Critics of design research argue that, at best, it can provide formative insights that must then be tested through more controlled experimentation (Barab, 2006). Critics argue further that design research is not a structured methodology but rather a loose collection of methods (Kelly, 2004), neither replicable, nor generalisable. These are legitimate claims, for design research is an emergent theoretical practice. That said, there is still a great deal we can learn from this perspective about recognizing the limitations of traditional methods and acknowledging the need for more integrative measures that are more successful at describing the impact of educational technology upon learning when situated in practice. In contrast to experimental studies, readily carried out with many participants in a controlled laboratory setting, research examining learning interactions most often requires intensive, fine-grained, high-frequency repeated assessment. These studies are time-consuming and difficult to carry out. As well, the nature of inquiry often demands that the researcher become part of the process. This further complicates matters, as traditionally such involvement would be seen as confounding the assessment process. However, within an evaluative framework that bases itself upon the premise that learning is a complex, dynamic, non-linear process, this involvement is seen as an inevitable, and therefore necessary element of inquiry (Jörg, Davis, and Nickmans, 2007). As Reeves (2007) notes, the advantage of such a research paradigm is that it invites collaboration between researchers and practitioners in the identification of teaching and learning challenges, and the creation, testing, and refinement of solutions. For too long we have developed and lab-tested innovative e-learning tools, which are subsequently inserted into the classroom without an adequate understanding of the context in which the tool is used.

4. Conclusion

Given the multimodal nature of interactive technology, it has a tremendous potential to support a variety of relationships and introduce new learning perspectives into students' understanding of complex subject matter. Examining the nature of these interactions and the way in which they foster depth of understanding is crucial to an appreciation of the role educational technology plays in learning. It demands an understanding of how to best support student learning in an integrated, holistic way, and how to leverage technology to support this process; which, in turn, demands of us that we develop evaluative tools capable of capturing the learning process that occurs when students interact with technology. Reeves (2007) has suggested that as educational technologists we may need to rethink our view of the field as a "science". Rather, if we accept that educational technology is first and foremost a design field then we can frame related inquiry with that perspective in mind. As a design field the goal of educational technology can shift from an experimental model to a more iterative model aimed at deriving design principles to inform future development and implementation of multimedia tools. From a learning perspective, by adopting an evaluative framework that takes a more flexible approach to measuring more meaningful learning effects associated with multimedia environments, we can examine the capacity of these environments to support more complex involvement with the learning material. We might then leverage technology to deepen understanding, by focussing less on knowledge outcomes and increasingly on the process by which understanding develops.

Acknowledgements

I am greatly indebted to my colleagues in Biomedical Communications at University of Toronto. In particular I would like to thank Professor Linda Wilson-Pauwels, former graduate student Diana Kryski, as well as Teddy Cameron (Discovery Commons) for their contributions to this paper. In addition I would like to acknowledge the significant contributions of my colleagues in Anatomy, Professor Mike Wiley, Chair of Anatomy, and Professor Emerita Patricia Stewart for their ongoing support and investment in research examining the impact of educational technology. Finally, I would like to thank Professor Earl Woodruff, Associate Chair of Human Development and Applied Psychology at the Ontario Institute for Studies in Education, my mentor in all things related to *understanding*.

References

Ainsworth, S. (2008) “How Should we Evaluate Multimedia Learning Environments?”, In J. Rouet, R. Lowe, and W. Schnotz (Eds.), *Understanding Multimedia Documents,* Springer Science + Business Media, LLC, New York, pp 249–265.

Anglin, G. J., Vaez, H., and Cunningham, K. L. (2004) “Visual Representations and Learning: the Role of Static and Animated Graphics”, In D.H. Johanssen (Ed.) *Handbook of Research on Educational Communications and Technology,* Lawrence Erlbaum Associates, Inc., New Jersey, pp 865–916.

Barab, S.A. (2006), “Design-based Research: A Methodological Toolkit for the Learning Scientist”, In K. Sawyer (Ed.) *The Cambridge Handbook of The Learning Sciences,* Cambridge University Press, Cambridge, pp 153–169.

Brown, A. L. (1992) “Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions”, *Journal of the Learning Sciences*, Vol 2, pp 141–178.

Clark, R. (1994) “Media Will Never Influence Learning”, *Educational Technology Research and Development*, Vol 42, No. 2, pp 21–29.

Confrey, J. (2006), “The Evolution of Design Studies as Methodology”, In K. Sawyer (Ed.) *The Cambridge Handbook of The Learning Sciences,* Cambridge University Press, Cambridge, pp 135–151.

Collins, A. (1992) “Toward a Design Science of Education”, In E. Scanlon and T. O’Shea (Eds.), *New directions in Educational Technology,* Springer-Verlag, New York, pp 15–22.

Collins, A., Joseph, D., and Bielaczyc, K. (2004) “Design Research: Theoretical and Methodological Issues”, *Journal of the Learning Sciences*, Vol 13, No. 1, pp15–42.

Visscher, K., Woolridge, N., Jenkinson, J., and Kain, K. “Effectiveness of Timeline Pacing as a Postproduction Cueing Strategy in an Educational Three-Dimensional Animation for Undergraduate Medical Students: A Randomized Controlled Trial”, *Journal of Biocommunication,* Vol 35, No. 1 pp E3–E8.

Garg, A.X., Norman, G.R., and Spero, L. (2002) “Is There any Real Virtue of Virtual Reality? The Minor Role of Multiple Orientations in Learning Anatomy from Computers”, *Journal of Academic Medicine*, Vol 77, pp S97–S99.

Hidrio, C. and Jamet, E. (2008) “Learning from a Multimedia Explanation: A Comparison of Static Pictures and Animation”, In J. Rouet, R. Lowe, and W. Schnotz (Eds.), *Understanding Multimedia Documents,* Springer Science + Business Media, LLC, New York, pp 103–120.

Jarodzka, H., Scheiter, K., Gerjets, P., and Van Gog, T. (2009) “In the Eyes of the Beholder: How Experts and Novices Interpret Dynamic Stimuli”, *Learning and Instruction*, in press.

Jenkinson, J., Stewart, P.A. and Cameron, T. (2007). [Animating Subcellular Events: Educational or Entertaining?]. Unpublished raw data.

Jörg, T., Davis, B., and Nickmans, G. (2007) "Towards a New, Complexity Science of Learning and Education", *Educational Research Review*, 145–156.

Kelly, E., A. (2004) “Design Research in Education: Yes, but is it Methodological?”, *Journal of the Learning Sciences*,Vol 13, No.1, pp 115–128.

Kryski, D. (2008) “Enhancing Medical Students’ Clinical Reasoning Ability Through Visualisation of Spatially Complex Gross Anatomy with a Computer-based, Three-dimensional Model of the Pterygopalatine Fossa”, MScBMC dissertation, University of Toronto.

Lowe, R. (2008) “Searching User-controllable Animations During Learning”, In J. Rouet, R. Lowe, and W. Schnotz (Eds.), *Understanding Multimedia Documents,* Springer Science + Business Media, LLC, New York, pp 149–167.

Lui, M., Jenkinson, J., Bacjar, J., Piquette-Miller, M., and Wilson-Pauwels, L. (2006) “Integrating Instructional Interactivity in a 2-D Animated Learning Environment to Enhance its Educative Value: An Online Pharmacogenetics Course for Practicing Pharmacists”, *Journal of Biocommunication*, Vol 32, No. 3.

Luursema, J.-M., Verwey, W.B., Kommers, P.A., Geelkerken, R.H. and Vos, H.J. (2006) “Optimizing Conditions for Computer-assisted Anatomical Learning”, *Interacting with Computers,* Vol 18, No. 5, pp 1123–1138.

Mayer, R. E., and Moreno, R. (1998) “A Split-attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory”, *Journal of Educational Psychology*, Vol 90, pp 312–320.

Mayer, R. E., and Moreno, R. (1999) “Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity”, *Journal of Educational Psychology*, Vol 91, pp 358–368.

Mayer, R. E., and Moreno, R. (2003) “Nine Ways to Reduce Cognitive Load in Multimedia Learning”, *Educational Psychologist*, Vol 38, pp 43–52.

Mayer, R.E. (2005) “Cognitive Theory of Multimedia Learning”, In: R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning,* Cambridge University Press, New York, pp 183–200.

Mayer, R.E., and Anderson, R.B. (1991) “Animations Need Narrations: an Experimental Test of a Dual-coding Hypothesis”, *Journal of Educational Psychology*, Vol. 83, pp 484–490.

Nicholson, D. T., Chalk,T., Funnell,W.R.J. and Daniel, S.J. (2006) “Can Virtual Reality Improve Anatomy Education? A Randomised Controlled Study of a Computer-generated Three-dimensional Anatomical ear Model”, *Medical Education*, Vol 40, No. 11, pp 1081–1087.

Oliver, M. (2000) “An Introduction to the Evaluation of Learning Technology”, *Educational Technology & Society*, Vol 3, No. 4, pp 20–30.

Paas, F., Ayres, P., and Pachman, M. (2008) “Assessment of Cognitive Load in Multimedia Learning”, In D.H. Robinson and G. Schraw (Eds.), Recent Innovations in Educational Technology That Facilitate Student Learning, Information Age Publishing Inc., Charlotte, NC, pp 11–35.

Paivio, A. (1983) "The Empirical Basis for Dual Coding", In J.C. Yuille (Ed.), *Imagery, Memory, and Cognition: Essays in Honour of Allan Paivio,* Erlbaum, New Jersey, pp 307–328.

Ploetzner, R., and Lowe, R. (2004) "Dynamic Visualisations and Learning", *Learning and Instruction*, Vol 14, pp 235–240.

Reeves, T.C. (2007), "Design Research From a Technology Perspective", In J. Van Den Akker et al. Eds. Educational Design Research, Routledge, New York, pp 52–66.

Robson, C. (2002), *Real World Research*, Blackwell Publishing, Oxford.

Siegler, R. S. and Crowley, K. (1991) "The Microgenetic Method: A Direct Means for Studying Cognitive Development", *American Psychologist*, Vol 46, pp 606–20.

Sweller, J., Van Merriënboer, J., and Paas, F. (1998) "Cognitive Architecture and Instructional Design", *Educational Psychology Review*, 10, 251–296.

Sweller, J. and Chandler, P. (1994) "Why Some Material is Difficult to Learn", *Cognition and Instruction*, 12, 185–233.

Sweller, J. (1988) "Cognitive Load During Problem Solving: Effects on Learning", *Cognitive Science*, Vol 12, 257–285.

Stewart, P., Tymianski, M., Jenkinson, J., Cameron, T., and Ballyk, D. (2008). [Toward Defining a Pedagogical Framework for the Development of Interactive Learning Tools]. Unpublished raw data.

Tasker, R. and Dalton, R. (2008) "Visualizing the Molecular World – Design, Evaluation, and use of Animations", In J.K Gilbert, M. Reiner, and M. Nakhlen (Eds.), *Visualisation: Theory and Practice in Science Education,* Springer Science + Business Media, LLC, New York, pp 103–131.

Van Gog, T., Scheiter, K.(2009) "Eye tracking as a Tool to Study and Enhance Multimedia Learning", *Learning and Instruction*, in press.

Van Gog, T., Paas, F., and van Merrienboer, J.G. (2005) "Uncovering Expertise-Related Differences in Troubleshooting Performance: Combining Eye Movement and Concurrent Verbal Protocol Data", *Applied Cognitive Psychology*, Vol 19, 205–221.

Vicente, K. (2004) *The Human Factor: Revolutionizing the Way People Live With Technology*, Routledge, New York, p 46.

Vicente, K. (1999) *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-based Work*, Lawrence Erlbaum and Associates, New Jersey.

