Introduction

External and internal representations in multimedia learning

New technologies in general, and multimedia in particular, play an increasingly important role in education. The term “multimedia” refers to the combination of multiple technical resources for the purpose of presenting information represented in multiple formats via multiple sensory modalities. Accordingly, multimedia resources can be considered at three different levels: the technical level refers to the technical devices (i.e. computers, networks, displays, etc.) that are the carriers of signs; the semiotic level refers to the representational format (i.e. texts, pictures, and sounds) of those signs; the sensory level refers to the sensory modality of sign reception (i.e. visual or auditory modality).

A number of misconceptions arise amongst educators because of a failure to distinguish these different levels. One misconception is found with those who treat multimedia primarily in terms of the information technology involved (and therefore primarily as a matter for computer scientists and engineers) and are generally unaware that they address only one of its levels. In fact, the other two levels are equally important and a proper understanding of them requires expertise in cognitive science, psychology, and educational science (rather than information technology). Another misconception also related to a failure to recognise the different levels of multimedia is that the technical medium itself is presumed to have an impact on learning. In fact, previous research on so-called “media-effects” has clearly established that it is misguided and overly simplistic to compare different technical media with regard to their effects on learning without taking account of semiotic and sensory aspects.

Rather than searching for technical media-effects, research on learning and instruction should focus on the semiotic and sensory levels of multimedia. The emphasis should be on the effects of different forms of external representation such as texts and graphics (either static or animated) on comprehension and learning. Unfortunately, there are also misconceptions with regard to multimedia’s semiotic and sensory levels. A widespread misconception of this type is the assumption that rich multimedia learning environments result in extensive cognitive processing and thus create elaborated knowledge structures. This view has meant that the possibilities of the latest information technologies are sometimes exploited to the limit in order to provide learners with multiple forms of representation that address a variety of different sensory modalities. The resulting learning materials include a diverse range of
components including animation, videoclips and various possibilities for interaction. However recent research indicates that multiple external representations and multiple modalities are not always beneficial for learning. For this reason, a better understanding of the processing demands associated with different kinds of representation and their function in comprehension and learning is required. It is becoming increasingly apparent that during multimedia learning, complex interactions take place between the external representations provided by increasingly sophisticated educational technologies and the internal (mental) representations constructed by learners whose instruction is presented via those technologies. The potential effectiveness of educational multimedia materials is therefore likely to be influenced by the extent to which their design takes these complex interactions into account.

The papers comprising this Special Issue aim to clarify the conditions under which multiple forms of representation are likely to support comprehension in multimedia learning environments and how specific features of multimedia can help learners to learn complex subject matter. The papers explore the requirements for achieving a suitable interplay between external and internal representation in order to foster effective learning from multimedia.

Mayer’s contribution focuses on multimedia presentations that contain spoken or written words combined with illustrations or animations that are designed to foster meaningful learning. In order to predict under which conditions learners profit more from multimedia messages than from verbal-only messages, Mayer presents a cognitive theory of multimedia learning that is based on three assumptions: the active learning assumption, the dual channel assumption, and the limited capacity assumption. According to this theory, meaningful learning requires that the learner engages in active cognitive processing. Active processing in turn requires that the learner pays attention to relevant words and pictures, organizes the corresponding information into coherent verbal and pictorial mental representations, and finally integrates verbal and pictorial representations with each other and with prior knowledge. Cognitive processing takes place in both a visual (pictorial) channel and an auditory (verbal) channel, each of which has a very limited capacity.

The theory presented by Mayer predicts that learners learn more deeply from words and pictures than from words alone (multimedia effect), that learners learn more deeply when extraneous material is excluded rather than included (coherence effect), and that they learn more deeply when printed words are placed near rather than far from corresponding pictures (spatial contiguity effect). Furthermore, he assumes that learners learn more deeply when words are presented in conversational rather than formal style (personalization effect). Practical implications of Mayer’s work are that instruction should present words and pictures rather than words alone, but should exclude extraneous words and pictures. Furthermore, corresponding words and pictures should be presented near each other, and the verbal component should use a conversational style.

Mayer’s findings are not confined to “high-tech” media. His theory and instructional principles apply for both book-based and computer-based learning environments. Although some forms of more advanced technology allow for instructional methods that are simply not possible with other technologies, the general principles
of instructional design are essentially the same irrespective of the type of media concerned. This is because the principles involved are not driven by technology, but rather derive from the functioning of the human mind and perceptual system.

Schnotz and Bannert propose a combined model of text and picture comprehension that contributes to a theory of multimedia learning. Like Mayer, these authors assume that learners actively select, organize and integrate verbal as well as pictorial information in a task-oriented way so as to construct mental representations that seem to be suited to anticipated demands. However, in contrast to Mayer’s model, the model developed by Schnotz and Bannert includes a basic distinction between descriptive and depictive representations. Descriptive representations result from symbol processing, whereas depictive representations result from analogical structure mapping. Because pictures permit a given piece of subject matter to be visualized in a variety of different ways, the model allows for the possibility that the specific form of visualization used in a picture can affect the structure of the resulting mental model and that model’s computational efficiency for specific tasks. The Schnotz and Bannert model predicts that adding pictures to a text may not always be beneficial for learning but rather may have negative effects if poorly matched to the learning task. This would mean that while task-appropriate pictures are likely to support learning, task-inappropriate pictures may actually interfere with the mental model construction processes that underpin effective learning. This assumption contradicts the traditional implication drawn from Paivio’s (1986) dual coding theory that adding pictures to a text is generally beneficial for learning.

Schnotz and Bannert report an experiment that investigated the effects of different external representations on mental model construction. Their empirical findings confirm the prediction of a possible detrimental effect resulting from task-inappropriate pictures. The study suggests that pictures facilitate learning if individuals have low prior knowledge and if the subject matter is visualized in a task-appropriate way. However, if individuals with higher prior knowledge receive pictures in which the subject matter is visualized in a task-inappropriate way, then these pictures may interfere with the construction of a task-appropriate mental model. In this respect, the findings contradict predictions commonly made on the basis of traditional dual coding theory.

Schnotz and Bannert’s findings emphasize the need for careful consideration of the type of visualization to be included when designing instructional material. Carefully considered visualizations are important not only for individuals with low prior knowledge who need pictorial support in constructing mental models. Well-designed pictures are also important for individuals with high prior knowledge because these individuals may be hindered in their mental model construction through inappropriate forms of visualization.

The next two papers are concerned with learning from animations, that is, dynamic depictions that can be used to make change processes explicit to the learner. Current educational practice and conventional instructional design wisdom indicate that animations are frequently assumed to be superior to static illustrations as tools for learning. In order to comprehend a dynamic situation that is externally represented by a static graphic, the learner must first construct a dynamic mental model by him/herself
from the static information provided. In contrast, animations can offer the learner a "ready-made" explicit dynamic representation of the situation. Thus, they appear on the surface to provide just the type of external support that would be required for the construction of a dynamic internal mental model. However, as indicated by the two contributions on animation, merely providing learners with the dynamic information in an explicit form does not necessarily result in better learning.

Lowe considers the influence that the distinctive display characteristics of animations have on the effectiveness of animated learning materials. He points out that animations can confront learners with additional and qualitatively different information processing demands from those they face with static graphics. It is suggested that the dynamic characteristics of animations are likely to be particularly crucial when these displays are used to represent complex subject matter for learners who are relative novices in the depicted domain. In order to build an effective mental model of a dynamic referent situation, information that learners extract from the external representation needs to be of central thematic relevance to that situation. With a complex animation, this involves selectively attending to those aspects that are particularly salient as far as the referent situation is concerned (rather than to those aspects that are readily noticeable simply because of their superficial appearance). However, novices are poorly equipped to identify these aspects within the rich flux of information in such animations because they lack the necessary domain-specific background knowledge. As a consequence, the information they extract is likely to be influence primarily by the extent to which it is perceptually conspicuous.

In contrast to many other investigations of learning from animation, the research reported by Lowe used animations that incorporated a high degree of user control. This approach was taken because one suggested reason for the lack of educational effectiveness of normal non-interactive animations is that they do not allow learners to determine the pace and direction of the presentation. Subjects in Lowe’s study used a complex animation to help them learn how to make predictions about the changes that would be expected in a graphic display over time. The findings suggest that, despite the opportunity to freely interrogate the animation, domain novices tended to neglect its subtle yet thematically relevant aspects. Instead, they extracted information about display components whose changes in position and form produced a marked perceptual contrast with their graphic context. A follow-up prediction task indicated that this perceptually conspicuous information (rather than information of thematic relevance) tended to be preferentially retained. A key implication from this research is that even animations allowing a high degree of user control may have to incorporate considerably more support and direction than is currently provided if they are to function as effective tools for learning.

Lewalter also questions the widespread assumption that animations result in better learning than static pictures and examines whether the two kinds of visual displays lead to different cognitive processing. She argues that the difference between their respective cognitive processing demands is twofold. On the one hand, directly supporting the construction of a dynamic mental model through an animation may reduce the load of cognitive processing. On the other hand, the transitory nature of dynamic
visuals may cause higher cognitive load because learners have less control of their speed of processing. A crucial question therefore is whether, and to what extent, learners really make use of the information provided by animations.

Accordingly, Lewalter focuses on cognitive and metacognitive strategies in learning from animations combined with expository texts. She shows that either adding animations or adding static illustrations can result in better learning about a physics topic (gravitation and optics). However, she found no difference between animations and static illustrations with respect to knowledge acquisition about facts, and only a small non-significant difference in favour of the animation group with respect to comprehension. Students who learn from animations seem to use different strategies from those used by students who learn from static illustrations. Whereas learners given static illustrations used rehearsal strategies more frequently and showed significantly more planning of their learning than subjects with animations, the latter made slightly more use of elaboration strategies. Whereas the use of rehearsal strategies affected fact learning but not comprehension, the use of metacognitive control strategies affected comprehension but not fact learning.

Lewalter’s results indicate that dynamic visuals are not generally superior to static visuals because learners with animations were not better than learners with static visuals with regard to recall and only marginally better with regard to comprehension. This provides confirmatory evidence for Lowe’s finding that explicit presentation of dynamic aspects of the content in multimedia learning environments does not necessarily have a positive impact on learning. In many cases, the use of static visuals including conventional signs for motion (like arrows) or the use of series of frames may be sufficient for successful understanding and learning. Further, it is possible that the potentially supportive effects of animations may be lost because individuals have less control of their processing in learning from animations and thus have restricted opportunities to use appropriate learning strategies.

Stern, Aprea and Ebner deal with the use of graphs as tools for presenting information, for reasoning, and for transfer. Active use of graphs is rarely learned systematically in schools, and accordingly, many individuals have difficulties in mapping graphical entities appropriately onto content entities, and spatial relations onto semantic relations. Stern et al. examine how adult learners use graphs as tools for reasoning and for the transfer of knowledge from one topic to another within a given knowledge domain. Their results suggest that learners in the domain of economics who actively constructed a linear graph with appropriate instructional guidance have higher transfer performance than learners who were simply presented a linear graph. Active creation of a graphical representation seems to be a powerful transfer tool, especially for learners with high prior knowledge. The higher the quality of the created graph, the higher is their transfer performance. But even learners with low prior knowledge can actively construct graphs if provided with sufficient help. Such active graph construction also turns out to be efficient for these less knowledgeable learners.

Stern, Aprea and Ebner argue that by actively constructing graphs based on text information, learners become more aware of the representational elements and relations as well as how these elements and relations have to be mapped onto the
specific content domain. This may result in more elaborated knowledge structures that are organized more clearly than those that result from simply reading ready-made graphs alone.

The last two articles in this Special Issue deal with the use of multiple external representations (i.e., those including more than two representational formats) to develop comprehension and learning. Kozma examines differences between chemistry experts and chemistry novices with regard to the use of representations such as chemical equations, graphs, molecular-level animations, and video segments showing lab experiments. These representations are characterized as material resources that support thinking and social interaction. The distinctive features of each of these types of representations imply cognitive and social affordances and constraints for the practice of inquiry and discourse when participants try to construct shared understanding of chemical phenomena. In both experimental and naturalistic studies, Kozma investigates how expert and novice scientists use the cognitive and social affordances that multiple representations make available. As the author points out, expert scientists use different representations for different purposes. They coordinate features within and across multiple representations to reason about their research. They also move easily across different representations and use them together to understand chemical phenomena. Further, they spontaneously create representations and coordinate multiple representations in order to argue, to explain, and to justify their views. Novices, on the contrary, have difficulties in moving across or connecting multiple representations. They also find it difficult to connect representations to the physical phenomena they stand for. Rather, their understanding and discourse are constrained by individual representations.

Kozma shows that the discourse among novices in the wet lab is somewhat different from their discourse in a computer lab with molecular modeling software. In the wet lab, the discourse focuses on chemical substances, equipment, and procedures. In the computer lab, the discourse focuses on molecular structures. In this respect, it has more similarity with the discourse of expert chemistry scientists. However, their discourse includes fewer references to the substances and phenomena that can be observed in the wet lab. In other words, there is both a lack of connection among representations and a lack of connection between representations and phenomena.

Kozma suggests several principles for the design and use of interactive multimedia learning environments that present coordinated multiple representations. Such technology-based systems should help learners to make the necessary connections and support collaborative efforts for shared comprehension of scientific phenomena.

Seufert addresses the question of how instruction can assist individuals to interconnect external multiple representations in order to foster mental coherence formation from these representations. Learners frequently fail to create referential connections between corresponding elements and corresponding relations of such multiple representations and thus do not make adequate use of these representations for mental coherence formation. Seufert investigates the effects of different kinds of instructional help on coherence formation from a set of different representations of a phenomenon from chemistry with learners of different prior knowledge. One group of subjects received directive help in the form of additional text paragraphs that
explained explicitly how the different representations could be mapped onto each other. Another group received non-directive help consisting of a list of additional sentences with hints intended to prime the learners’ search for corresponding elements and corresponding relations in different representations. A control group received the learning material with no instructional help.

The results of Seufert’s study indicate the presence of a complex interplay between the fluency of cognitive processing, higher order coherence formation, amount of memorization, and the individual’s aims. As would be expected from a repetition of information under the conditions of fluent cognitive processing, directive help turned out to be more effective for recall than non-directive help. However, non-directive help on the one hand stimulated higher order coherence formation, but on the other hand resulted in low fluency of processing. Seufert’s interpretation of this finding is that the visual presentation of non-directive help required the learner to interrupt semantic processing repeatedly and that these interruptions interfered with mental encoding. Surprisingly, learners with low knowledge showed highest recall when they had received no help. Instructional help (either directive or non-directive) seemed to have overtaxed the cognitive capabilities of these learners, whereas no help allowed them to intentionally concentrate on memorization according to a simple recall strategy.

With regard to comprehension, both directive and non-directive help were found to be effective for learners with medium prior knowledge. While it seems that learners with low prior knowledge lack the cognitive requirements to make effective use of help, learners with high prior knowledge are unlikely to require the help anyway. Overall, both directive and non-directive help have the potential to facilitate comprehension, provided that learners are capable of using this help and are actually in need of such assistance.

We hope that this Special Issue will contribute to a better understanding of the complex interplay between external representations and internal (mental) representations involved in learning from multimedia. In addition, we trust that it will stimulate further investigations in this rapidly expanding area of research that has such important implications for future educational practice.

We express our special thanks to Susan Goldman and Peter Reimann who not only prepared most insightful reviews of the papers for this Special Issue, but also provided illuminating commentaries on the papers and valuable suggestions about promising directions for further research.

Wolfgang Schnotz
Richard Lowe

a University of Koblenz-Landau, Department of General and Educational Psychology, Thomas-Nast-Str.44, D-76829 Landau, Germany

b Faculty of Education, Curtin University, GPO Box U1987, Perth, WA, 6845, Australia