# Does instructional format really matter? Cognitive load theory, multimedia and teaching English literature

**Stewart Martin, Faculty of Education, University of Hull**

## Abstract

This article reports a quasi-experimental study on the effects of multimedia teaching and learning in English Literature – a subject which places high cognitive load on students. A large-scale study was conducted in 4 high-achieving secondary schools to examine the differences made to students' learning and performance by the use of multimedia and to relate this to different kinds of multimedia. Statistical significance and effect size calculations indicated that the equivalent of one grade level in General Certificate of Secondary Education (GCSE) was associated with the use of advanced and integrated multimedia, and that this was stronger than the effects of schools and sex of the students. It was found that advanced multimedia software eased cognitive overload, particularly in the area of intrinsic cognitive load. Limitations of the study are drawn, including the needs to examine process variables and learner-related variables. Conclusions and implications for further research and for enhancing teaching and learning with multimedia are made.

Keywords: multimedia, cognitive load, effective education, pedagogy, English Literature,

## Notes on contributor

Stewart Martin is Senior Lecturer and director of postgraduate taught programmes in the Faculty of Education, University of Hull. His research is funded by the European Commission, the UK Teacher Development Agency, the Higher Education Funding Council, the British Council (Digital Economy), and the EPSRC. His research interests include the educational application of digital technology, citizenship and images of the self, educational leadership, and organisational change. He is currently Chair of the International Virtual Environments Research Group (iVERG) http://www.iverg.com and a member of the international editorial board for *Educational Research and Evaluation*.

## Introduction

The use of multimedia is argued to have the potential to significantly improve instructional efficacy, particularly with regard to the successful learning of information and the development of understanding (Mayer, 2008 Mayer, R.E. 2008. Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63: 760–769. ; Miller, Chang, Wang, Beier, & Klisch, 2011 Miller, L.M., Chang, C.I., Wang, S., Beier, M.E. and Klisch, Y. 2011. Learning and motivational impacts of a multimedia science game. *Computers & Education*, 57: 1425–1433. ), but concerns persist about the degree to which its design and use have realised or optimised such potential (Argyris,1976; Massa & Mayer, 2006 Massa, L.J. and Mayer, R.E. 2006. Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style?. *Learning and Individual Differences*, 16: 321–335. ; Schnotz & Kürschner, 2007 Schnotz, W. and Kürschner, C. 2007. A reconsideration of cognitive load theory. *Educational Psychology Review*, 19: 469–508. ; Sweller & Chandler, 1994 Sweller, J. and Chandler, P. 1994. Why some material is difficult to learn. *Cognition and Instruction*, 12: 185–233. ; Tabbers, Martens, & Van Merriënboer, 2000 Tabbers, H.K. and Martens Van Merriënboer, R.L. J.J.G. *Multimedia instructions and cognitive load theory: Split-attention and modality effects*. Paper presented at the National Convention of the association for Educational Communications and Technology. February. ). The application of cognitive load theory (CLT) (Sweller, Van Merriënboer, & Paas, 1998 Sweller, J., Van Merriënboer, J.J.G. and Paas, F. 1998. Cognitive architecture and instructional design. *Educational Psychology Review*, 10: 251–296. ) has been at the forefront of much experimental work in this area, and it articulates important implications for optimising the design of educational multimedia (Mayer, 2003 Mayer, R.E. 2003. The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, 13: 125–139. , 2009 Mayer, R.E. 2009. *Multimedia learning*, Cambridge, , UK: Cambridge University Press. ; Mayer & Moreno, 2002 Mayer, R.E. and Moreno, R. 2002. Aids to computer-based multimedia learning. *Learning and Instruction*, 12: 107–119. ).

This article reports an experimental empirical study of the use of multimedia in the teaching of English Literature and the effect of its use on scores from a measure of knowledge and understanding of Shakespeare's *Macbeth*.

## Cognitive load theory

Cognitive load theory (CLT) seeks to explain why some material is more difficult to learn than other material (Chandler & Sweller, 1991 Chandler, P. and Sweller, J. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8: 293–332. ). It is based on the proposition that the human brain uses two types of memory: short-term and long-term memory, where short-term memory is seen as having limited storage capacity and long-term memory is seen as having almost unlimited storage capacity (Sweller, 1994 Sweller, J. and Chandler, P. 1994. Why some material is difficult to learn. *Cognition and Instruction*, 12: 185–233. ). CLT proposes that the existence of these two types of memory in humans is important because it determines and has been determined by the way we learn. Using short-term memory, we develop *schema* (e.g., cognitive constructs, organised knowledge, or classifications of problems into categories: “cognitive constructs that incorporate multiple elements of information into a single element with a specific function” (Paas, Renkel, & Sweller, 2003 Paas, R., Renkel, A. and Sweller, J. 2003. Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38: 1–3. , p. 2)) and

store these in long-term memory. Schema help us with tasks such as solving problems that we have not seen before by using our learning about similar kinds of problems we have solved in the past. With practice, using our schema can speed up problem solving and task execution by partially automating our cognitive activity when responding to situations or problems that are similar to ones we have learned about in the past. We use the limited capacity of short-term memory to manipulate existing schema (or to create new ones) and apply these to the solution of problems which would otherwise prove too complex for us to deal with if we always had to begin from first principles.

CLT argues that whether some material is easier to learn than other material depends in large part on the degree to which we are able to reduce the amount of processing (cognitive load) needed to solve a problem or learn something new by using schema acquisition and automation: "schemas effectively increase the amountof information that can be held in working memory by chunking individual elements into a single element" (Sweller, 1994 Sweller, J. and Chandler, P. 1994. Why some material is difficult to learn. *Cognition and Instruction*, 12: 185–233. , p. 299). "Chunking" is thought to be a common feature of the way short-term memory operates to organise information when constructing schema (Chase & Simon, 1973 Chase, W.G. and Simon, H.A. 1973. Perception in chess. *Cognitive Psychology*, 4: 55–81. ). Neuroscience speculates that cognitive load may be directly linked to physical "memory load" as a result of the way the brain coordinates the firing of neurons, but this phenomena is not well understood as yet and alternative mechanisms have also been proposed to explain how chunking may be handled in working memory (O'Reilly, Busby, & Soto, 2003 O'Reilly, R.C., Busby, R.S. and Soto, R. 2003. "Three forms of binding and their neural substrates: Alternatives to temporal synchrony". In *The unity of consciousness: Binding, integration, and dissociation*, Edited by: Cleeremans, A. 168–190. Oxford, , UK: Oxford University Press. ).

CLT identifies a number of key factors which, independently, can increase the cognitive load on a learner (or can produce unhelpful types of cognitive load), when dealing with inherently complex material or when confronted by poor instructional design. Given the arguments that learning with multimedia can be more efficient and effective, a consideration of both of these latter factors is important.

In contemporary CLT, three expressions of cognitive load are understood to operate. *Extraneous cognitive load* is the difficulty, or load on the learner's working memory, associated with the design of instructional materials and the way these present information to the learner. High extraneous cognitive load is harmful to learning and is created as a result of unnecessary processing caused by the instructional design. *Germane cognitive load* is the load that is directed towards constructing, processing, and automating schemas. It can also be manipulated by the instructional design but is helpful to learning because it results from features of the design which direct attention towards relevant learning processes. *Intrinsic cognitive load* is directly attributable to the inherent complexity or difficulty of the material to be learned and may not be changed by the teacher; it is assumed to be unaffected by the instructional design and to be the product of a combination of the learner's prior knowledge and the intrinsic complexity of the learning material (Sweller & Chandler, 1994 Sweller, J. and Chandler, P. 1994. Why some material is difficult to learn. *Cognition and Instruction*, 12: 185–233. ).

Researchers in the field of CLT attempt to arrange the instructional control of cognitive load so as to optimise the load experienced by subjects in learning situations. The distinction

between optimising, as opposed to maximising or minimising, is important and is commonly misunderstood. The aim of cognitive load researchers is not simply to reduce the overall cognitive load on learners, as is sometimes incorrectly assumed, but to avoid extreme situations where there is too little load or too much load, because learning deteriorates in both situations (Young & Stanton, 2002 Young, M.S. and Stanton, N.A. 2002. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, 44: 365–375. ). Cognitive load researchers wish to produce both the optimum amount of load for learning and to promote load *of the right kind*; that is, they seek to optimise the load that contributes to learning (i.e., germane load) and reduce the load imposed by elements that hamper learning (i.e., extraneous load).

When learners find instructional tasks easy (e.g., when intrinsic load is low), any extraneous cognitive load imposed by the learning resources or context may have little or no significant negative effect on learning. This is not the case when tasks are more difficult and the intrinsic cognitive load is high; under these circumstances, it is important to take account of (and if possible reduce) the extraneous load on learners (Van Merriënboer & Sweller, 2005 Van Merriënboer, J.J.G. and Sweller, J. 2005. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17: 147–177. ).

However, in some learning situations it may be difficult to reduce the intrinsic load on learners because the learning tasks may be very complex, they may have unavoidably high-element interactivity, or may require the development and application of many different schema, such as in situations where multiple choices are available to the learner regarding the information to be selected and applied. Thiswould be the case in the example given below concerning English Literature, where evidence and judgement about the personality and interactions of a dramatic character are being selected and combined.

Research using CLT has sought to find ways to manage high intrinsic cognitive load (Pollock, Chandler, & Sweller, 2002 Pollock, E., Chandler, P. and Sweller, J. 2002. Assimilating complex information. *Learning and Instruction*, 12: 61–86. ) by approaches that take account of learner's prior knowledge (see Kalyuga, Ayres, Chandler, & Sweller, 2003 Kalyuga, S., Ayres, P., Chandler, P. and Sweller, J. 2003. The expertise reversal effect. *Educational Psychologist*, 38: 23–32. ) or that allow for the level of germane load imposed on learners by different instructional materials (Cierniak, Scheiter, & Gerjets, 2009 Cierniak, G., Scheiter, K. and Gerjets, P. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?. *Computers in Human Behavior*, 25: 315–324. ; Kalyuga, Chandler, & Sweller, 1998 Sweller, J., Van Merriënboer, J.J.G. and Paas, F. 1998. Cognitive architecture and instructional design. *Educational Psychology Review*, 10: 251–296. ; Salomon, 1984 Salomon, G. 1984. Television is “easy” and print is “tough”: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76: 647–658. ).

When considering the design of resources that teachers may use to promote learning, CLT argues that the physical integration of multiple sources of information is generally beneficial for learners. Physical integration happens when, for example, text and images are combined in multimedia applications or on the page of a textbook so that each does not simply replicate the content contained in the other. Physical integration eliminates the need for learners to split their attention between (for example) physically separate illustrations and text on a page or screen when learning material. Where physical integration has not happened, the learner's

attention is divided unhelpfully between the separate elements, as they attempt to process each one individually and make cognitive associations between them. This *split-attention effect* is regarded as unhelpful for learning because it increases extraneous load, and so learning materials featuring split-attention may overwhelm working memory capacity (Chandler & Sweller, 1992 Chandler, P. and Sweller, J. 1992. The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62: 233–246. ; Sweller, 1994 Sweller, J. and Chandler, P. 1994. Why some material is difficult to learn. *Cognition and Instruction*, 12: 185–233. ).

However, subsequent studies have found that in any given subject domain, certain learning resources which are beneficial for less expert learners become disadvantageous as learners become more expert (Kalyuga et al., 1998 Kalyuga, S., Chandler, P. and Sweller, J. 1998. Levels of expertise and instructional design. *Human Factors*, 40: 1–17. ). Inparticular, the physical integration of information as a means to minimise the split-attention effect becomes less helpful to learners as their expertise grows, and it becomes counter-productive for learning as expertise increases still further (Kalyuga et al., 1998 Kalyuga, S., Chandler, P. and Sweller, J. 1998. Levels of expertise and instructional design. *Human Factors*, 40: 1–17. ).

For more expert learners, the physical separation of information can be more advantageous than its integration, because they are likely to already possess the schema that the learning resources are attempting to promote in less experienced learners. As a result, learning resources may become subject, therefore, to an *expertise-reversal effect*. The expertise-reversal effect appears when more expert learners find it easier to handle complex instructional material but more difficult to learn from material that is designed to integrate separate elements in order to aid less experienced learners to construct appropriate mental representations (schema); in such cases, experienced learners are confronted with instructional guidance that is redundant for them, and this can be difficult to ignore, thus increasing cognitive load and reducing the efficiency of their learning (Kalyuga et al., 2003 Kalyuga, S., Ayres, P., Chandler, P. and Sweller, J. 2003. The expertise reversal effect. *Educational Psychologist*, 38: 23–32. ). Intrinsic cognitive load can therefore be determined only in the light of reference to a particular level of expertise (Schnotz & Kürschner, 2007 Schnotz, W. and Kürschner, C. 2007. A reconsideration of cognitive load theory. *Educational Psychology Review*, 19: 469–508. ).

Obtaining measures of individual cognitive load from learners can be problematic, not least because learners, faced with a new topic or domain, may find it hard to know whether any difficulty they experience is due more to the content or to the instructional design. In such circumstances, it can be difficult or impossible to identify reliably and disentangle the origins of extraneous and intrinsic cognitive load (Cierniak et al., 2009 Cierniak, G., Scheiter, K. and Gerjets, P. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?. *Computers in Human Behavior*, 25: 315–324. ). As a result of either one or the other being higher or lower for different learners, the *overall* cognitive load for different learners may be equally high. However, being able to take due account of the intrinsic load of learning tasks when using instructional multimedia is of importance for educators who are interested in managing such resources to obtain the greatest learning gains for individuals.

Because intrinsic load varies not just as a result of the inherent complexity of the learning material but also with the expertise of the individual learner in that subject or content area,

establishing the intrinsic load for individual learners is important for maximising their learning. However, the measurement and management of intrinsic cognitive load can be problematical, not least because objective measures are difficult to obtain and subjective measures, whilst easier to obtain, are not without their own difficulties.

One of the main problems with subjective (i.e., self-report) instruments in general is attributable to what Argyris described as the difference between *espoused theory* and *theory in use* (Argyris, 1976 Argyris, C. 1976. Theories of action that inhibit individual learning. *American Psychologist*, 31: 638–654. ), that is, the difference to be found between what individuals say they do and what they actually do. Learners may identify a particular approach to, or difficulty with, learning as the one they most frequently use or experience, but unless this is verified experimentally or by other means, we are unable to determine the accuracy of such reports, and a learner may well employ entirely different strategies (or experience entirely different difficulties) in practice from those they consistently report in good faith on questionnaires or during interview. This problematic limitation of instruments employing self-reporting (see Veenman, Prins, & Verheij, 2003 Veenman, M.V.J., Prins, F.J. and Verheij, J. 2003. Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology*, 73: 357–372. ) has also been identified with regard to the use ofinstruments proposed for measuring cognitive load. This becomes particularly problematic when cognitive load varies as a result of the learner's changing framework of reference and increased schema acquisition in response to the course of learning (i.e., as learner expertise increases), because the difficulties that are perceived by the learner and the associated degree of helpfulness of particular resources may be continuously changing as learning proceeds (Schnotz & Kürschner, 2007 Schnotz, W. and Kürschner, C. 2007. A reconsideration of cognitive load theory. *Educational Psychology Review*, 19: 469–508. ).

Research on multimedia and learning has demonstrated learning improvements in areas such as science (Moreno & Mayer, 1999 Moreno, R. and Mayer, R.E. 1999. Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91: 358–368. ; Zheng, Yang, Garcia, & McCadden, 2008 Zheng, R.Z., Yang, W., Garcia, D. and McCadden, E.P. 2008. Effects of multimedia and schema induced analogical reasoning on science learning. *Journal of Computer Assisted Learning*, 24: 474–482. ), management (Passerini, 2007 Passerini, K. 2007. Performance and behavioral outcomes in technology-supported learning: The role of interactive multimedia. *Journal of Educational Multimedia and Hypermedia*, 16: 183–211. ), chemistry (Lee, 2007 Lee, H. 2007. Instructional design of web-based simulations for learners with different levels of spatial ability. *Instructional Science*, 35: 467–479. ; Su, 2008 Su, K.D. 2008. An informative study of integrating multimedia technology into problem-solving for promoting students' abilities in general chemistry. *International Journal of Instructional Media*, 35: 339–353. ), physical education (Vernadakis, Avgerinos, Zetou, Giannousi, & Kioumourtzoglou, 2006 Vernadakis, N., Avgerinos, A., Zetou, E., Giannousi, M. and Kioumourtzoglou, E. 2006. Learning with multimedia technology – A promise or reality?. *Inquiries in Sport & Physical Education*, 4: 326–340. ), audio engineering (Cochrane, 2007 Cochrane, T. 2007. Developing interactive multimedia learning objects using QuickTime. *Computers in Human Behavior*, 23: 2596–2640. ), history (Williams, 2009 Williams, T. 2009. Multimedia learning gets medieval. *Pedagogy*, 9: 77–95. ), and physics (Stelzer, Gladding, Mestre, & Brookes, 2009 Stelzer, T., Gladding, G., Mestre, J.P. and Brookes, D.T. 2009. Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content. *American Journal of Physics*, 77: 184–190. ). The present study turns its attention to the little-explored area of multimedia use in learning and teaching in English Literature within

the UK secondary school curriculum, but the article argues that its ramifications extend to other contexts.

**Cognitive load theory and English Literature teaching and learning**

Teachers of English Literature in the UK generally seek to encourage the growth of a number of broad skills and abilities in their students as part of their preparation for public examinations such as the Standard Assessment Tests (SATs), the General Certificate of Secondary Education (GCSE), or Advanced Level studies (GCSE Advanced Levels), whilst also taking into account the requirements of the relevant examination board, as set out in its subject syllabus. Desired outcomes for dramatic works or novels, for example, are likely to require students to demonstrate:

- Recall of a sequence of events (the story or plot) in their correct sequential order;
- Understanding of the structure of the narrative;
- Recall of the names of and family relationships between characters;
- Insight into the individual characteristics and traits of characters;
- Knowledge of the dramatic relationships between characters;
- Awareness of the main themes and images present in the work;
- Recognition and understanding of the dramatic structure of the text;
- Understanding of how dramatic elements (e.g., themes and imagery) interact to create meaning and reveal insight;
- Understanding of how each dramatic character is developed;
- Understanding of historical or social influences;
- Understanding of the human condition including spiritual, moral, ethical social and cultural issues.

(Assessment and Qualifications Alliance [AQA], 2011a; Edexcel, 2010 Edexcel. 2010. *GCSE in English Literature: Scheme of work – Units 1 and 3* Retrieved from http://www.edexcel.com/Pages/Home.aspx)

In order to achieve high levels of performance in public examinations, regulatory bodies expect students to master and display knowledge, understanding, insight, andappreciation of this kind (AQA, 2008, 2009; Edexcel, 2011 Edexcel. 2011. *Examiner's report: June 2011 – Unit 3 Shakespeare and contemporary drama. Edexcel GCSE in English Literature (2ET01) Accreditied Specification*, Nottingham, , UK: Edexcel Publications. Retrieved from http://www.edexcel.com/quals/gcse/gcse10/english/literature/Pages/default.aspx). To facilitate the development of such learning, teachers commonly draw upon personal enthusiasm and make use of their instructional skill to encourage the high levels of student motivation and engagement which are normally necessary to develop the close familiarity with the text and the critical thinking and analytical skills which are needed. As part of this, teachers often make use of a range of techniques and resources to encourage and promote learning.

Despite such efforts, students often find the study of English Literature difficult because of the inherent complexity of the subject matter and the ways in which this is expressed, where many elements may be closely interrelated and because, as part of their preparation for

examinations, they often have to be able to apply their learning to alternative contexts, such as when discussing the contribution of one character or element of a literary or dramatic work to the other specific elements or structural features of that work, or to those of a different literary work, for example, one created by the same author or by a different author writing about a similar theme or context (Council for the Curriculum Examinations and Assessment [CEA], 2010, 2011; Edexcel 2008 Edexcel. 2008. *Examiner's report – June 2008: English Literature (1213)* Retrieved from http://www.edexcel.com/quals/gcse/gcse-leg/english/1213/Pages/default.aspx, 2009 Edexcel. 2009. *Examiner's report – June 2009: English Literature (1213)* Retrieved from http://www.edexcel.com/quals/gcse/gcse-leg/english/1213/Pages/default.aspx, 2011 Edexcel. 2011. *Examiner's report: June 2011 – Unit 3 Shakespeare and contemporary drama. Edexcel GCSE in English Literature (2ET01) Accreditied Specification*, Nottingham, , UK: Edexcel Publications. Retrieved from http://www.edexcel.com/quals/gcse/gcse10/english/literature/Pages/default.aspx). Achieving high examination results in English Literature requires students to display knowledge, understanding, and insight that can be gained only by effectively memorising, analysing, and synthesising many pieces of information, that is, many discrete interacting elements (AQA, 2011a; Edexcel, 2010 Edexcel. 2010. *GCSE in English Literature: Scheme of work – Units 1 and 3* Retrieved from http://www.edexcel.com/Pages/Home.aspx).

According to CLT, the processing of information in human cognition takes place within a limited working memory (Baddeley, 1992 Baddeley, A. 1992. Working memory. *Science*, 255: 556–559. ), and so there exists the physical capacity to handle only a very limited number of novel interacting elements at one time, possibly as few as two or three (Paas et al., 2003 Paas, R., Renkel, A. and Sweller, J. 2003. Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38: 1–3. ). Different information varies across a wide spectrum, from high-element interactivity to low-element interactivity, depending on the subject and the level of complexity of the actual content. The importance of this for learning is that whilst an element of low-interactivity material can be learned and understood individually (because it does not require consideration of other elements), this is not the case for high-interactivity material such as that commonly found in the subject matter of English Literature courses.

Learning the names and activities of the characters in a novel or play provides an example of low-element interactivity because each can be learned and understood without reference to any other items. This task imposes a relatively small cognitive load and may be handled easily. By contrast, learning how to describe the contribution of any of these characters to the dramatic structure and meaning of the work as a whole provides an example of high-element interactivity. Similarly, arguing how one character may be seen in a positive or negative light cannot be done independently of other characters and events in the dramatic work, because reaching a conclusion on this requires knowledge and understanding of events, ideas, and other characters and of how these all relate to each other; that is, the elements required for consideration interact.

The separate elements of high-interactivity material can be learned individually, but understanding them requires all of them and their interactions to be processed simultaneously. This can make high-element interactivity material (such as is found in many highly regarded works of literature) difficult to understand, because it can impose a cognitive load that exceeds the processing capacity of working memory (Ginns, 2006 Ginns, P. 2006. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16: 511–525. ).

However, whilst such material may be more difficult to learn, it is clearly not impossible, as evidenced by the successful learning that teachers of English Literature see in students on a regular basis. The way in which human cognition research allows for the learning of high-element interactivity material is by way of long-term memory, which is an extensive store containing large numbers of *schemas*. The store of schemas is subject to addition or revision in the light of further learning or new experiences (Bartlett, 1932 Bartlett, F.C. 1932. *Remembering*, Cambridge, , UK: Cambridge University Press. ; Neisser, 1967 Neisser, U. 1967. *Cognitive psychology*, London, , UK: Prentice Hall. ; Piaget, 1985 Piaget, J. 1985. *The equilibration of cognitive structures*, Chicago, IL: University of Chicago Press. ). A schema may consist of a large number of interacting elements which, if they each had to be processed separately, might easily exceed the capacity of working memory.

In the case of the literary illustration above, one example of a schema might be about behaviour, where a number of elements are combined to produce a classification of a character's actions as "negative" or "positive". Schemas are hierarchical, domain-specific knowledge structures that categorise multiple elements of related information as a single higher order element. In this, they differ from "constructs" (as in construct theory) because they: (a) are cognitive rather than emotional or ethical, (b) do not necessarily represent or articulate our core values, (c)do not focus on our key personal relationships, and (d) are relatively easy to modify or change (Kelly, 1995 Kelly, G.A. 1995. *The psychology of personal constructs Volume 1: A theory of personality; Volume 2: Clinical diagnosis and psychotherapy*, New York, NY: Norton. ). Controlling the use of schemas requires conscious effort on the part of the learner, but this can reduce with practice to the point where using a schema can become an automatic process rather than a controlled one.

In the present example, the elements being considered about the behaviour of a dramatic persona may include, selectively, characteristics or relationships that are seen as being either "bad" or "good", "kind" or "unkind", "generous", or "mean" and so on. Schemas for these characteristics and relationships can be retrieved from long-term memory and used in short-term memory, meaning that only a small number of elements have to be processed (used) in memory and the use of schemas may become automated so that they can be processed unconsciously to reduce further the load on working memory. According to CLT, it is by these processes that complex material can be handled by human cognitive architecture when it appears toexceed the capacity of working memory.

**The purpose of the research**

This article reports an experimental empirical study of the use of multimedia in the teaching of English Literature and the effect of its use on scores from a measure of knowledge and understanding of Shakespeare's *Macbeth*. The study was designed to explore the effect of multimedia use on learning using matched groups of students and teachers in four secondary schools and of differing formats of instructional material in multimedia and non-multimedia formats.

This study draws on previous work in which differences in expertise were found to give the largest and most reliable explanation for differences in performance between individuals. This study also took account of Kalyuga et al. (1998 Kalyuga, S., Chandler, P. and Sweller, J. 1998. Levels of expertise and instructional design. *Human Factors*, 40: 1–17. ), who found that subjective ratings of mental effort, that is, the mental effort associated with learning the instructional materials, was an effective and accurate proxy measure of cognitive load. The

present study follows Kalyuga and adopts the use of subjective ratings of mental effort with exemplar learning materials as an ecologically valid and reliable proxy for prior learning (and therefore for a main element of intrinsic cognitive load) when comparing the learning gains of groups which were using either existing or alternative (multimedia-based) approaches to learning, allowing for different levels of cognitive load attributable to the instructional materials.

The present study sought to explore the relationship between the intrinsic cognitive load imposed by different formats of learning resource and the effects of their use on learning gains for individuals in GCSE English Literature classes, in particular the differences made by, and the effects of, using multimedia in teaching and learning the GCSE English Literature syllabus (see below).

Experimental and control groups were compared to explore the degree to which the use or non-use of multimedia resources mediated these gains for learners with similar subject expertise when learning complex and demanding content. The intervention (the multimedia resources) was designed deliberately to reduce the cognitive load on students and was targeted to ease the cognitive load presented by complex material and its associated overload of short-term memory.

## The research procedures

### Sampling

From the group of 45 secondary schools used by a university in the North-East of England, UK, for placing trainee teachers during the school-based phases of their programme, 19 were selected because, within the previous 18 months, they had been subject to government inspections which in their resulting reports identified them as “good” or ‘outstanding”.

The designations “good” and “outstanding” by government inspectors carry significant currency within the UK; they are widely used (if contentious) government-endorsed judgements of the quality of educational provision and student performance which lead to associated benefits, including less intrusive inspection in the future. Inspection reports are published online, so praiseworthy inspections are useful to schools in promotional publicity, such as when seeking to recruit the most able and skilled faculty and also to parents when selecting the school they wish their children to attend. As the level of a state school's funding is closely linked to the number of pupils on its roll, increased competition for places at those which are perceived to be most successful ensures that such schools receive maximum income and acquire high status within their community. Inspection judgements are also referenced by the internet search engines of estate agents (realtors) and have a strong influence on the prices that can be commanded for residential property closest to schools judged to be very successful. There is a powerful incentive, therefore, for schools and individual teachers to strive for positive outcomes from government inspections.

Within the group of schools identified, a subset of nine had in common that in their reports their inspection teams had noted that faculty were highly skilled at identifying students' individual learning needs and in meeting them with a range of different, effective, individually targeted teaching and learning resources and strategies. In many other respects, set out below, the schools were broadly similar, despite serving communities providing student intakes of differing average levels of ability.

Inspectors noted in each school that teachers knew and taught their pupils well, had high expectations of them, matched classroom activity well to their needs, succeeded in having pupils make effective use of a wide range of research skills, were adventurous and imaginative in their teaching, and used a variety of teaching and learning styles effectively. Inspectors also commented that achievement in external examinations was good and that teachers knew how to match their teaching styles and learning materials effectively to pupils' different requirements.

From this subset of nine schools, four were selected for the present study on the grounds that they had well established English faculty who made relatively limited use of information technology in their approaches to learning and teaching. Technology use by the English faculty of these schools was largely confined to the use of Microsoft Office applications for word processing (e.g., for the production of student assignments) and for information retrieval, where teachers would access syllabus information from government websites or where, under their direction, students would print lesson support material (e.g., worksheets) from online repositories. None of the English departments in these schools made use of other computer resources or multimedia applications to support the learning of students or the teaching of faculty staff.

Participants in these four schools were divided into those replacing some of their normal English lessons with sessions where they used multimedia (the experimental group) and those having “normal” (non-multimedia) lessons (the control group); in School 1, each of these groups consisted of 133 students, in School 2 there were 100 students, in School 3 there were 102 students, and in School 4 there were 59 students (Table 1).

Existing teaching groups in the four schools were used to create a more natural teaching and learning setting for the half-term intervention than the strictly randomised groups that might feature in a fully experimental study; this also eased school timetabling difficulties. The study also adopted a quasi-experimental design, with equal numbers of experimental and control groups (classes) for each teacher. These groups were used to explore the effect on learning of multimedia use in teaching the GCSE English Literature syllabus (see AQA, 2008, 2009; Edexcel, 2011 Edexcel. 2011. *Examiner's report: June 2011 – Unit 3 Shakespeare and contemporary drama. Edexcel GCSE in English Literature (2ET01) Accreditied Specification*, Nottingham, , UK: Edexcel Publications. Retrieved from http://www.edexcel.com/quals/gcse/gcse10/english/literature/Pages/default.aspx; Qualifications and Curriculum Development Agency [QCDA], 2012).

**Controls**

Equal numbers of teaching groups and participants came from a number of existing classes in these schools, such that for any given subject teacher, the same number of classes they taught appeared in the multimedia (experimental) and non-multimedia (control) group (Table 1).

It is generally not possible to create full experimental conditions in field studies such as that reported here, mainly for practical reasons of securing the willing participation of schools and also for ethical reasons, such as if considering administering an intervention thought to be unhelpful or of no benefit. However, a number of elements could be controlled reasonably well: Experimental and control groups were matched for age-range and performance to date

to give a representation of “more able” and “less able” students (as determined by the school's assessment and “setting” arrangements) and as far as possible for numbers of males and females (except in School 3, which was all-female).

Within these constraints, classes were randomly selected for participation as far as possible (within the limits of the number of classes available in each school). Classes from each school were involved in the study for the same number of weeks (generally a complete half-term), for whole lessons at a time (around one hour in the case of each school), and in terms of the examination being studied for and the syllabus being followed. Teachers were all established in their school and very (but not identically) experienced in teaching their subject to the relevant examination level (GCSE) and were asked to make no changes to their existing teaching strategies and the resources they customarily used, except for the occasions when the experimental groups used multimedia resources. No controls were imposed for teacher gender, differences in teacher attitude towards multimedia or familiarity with computer use, and comparative measures of teaching strategy when using or not using multimedia were not undertaken.

The experimental groups used a specially developed multimedia application in their English Literature classes for approximately one hour (one lesson from their allocated two) per week over 10 weeks. This application covered the syllabus-relevant content for the text being studied. The control groups studied the identical content to the experimental groups during their English Literature classes (and in the same sequence and time period), with the same teacher, but they experienced only the teacher's usual approaches to teaching and learning, and they used the learning resources their teacher normally employed.

Each experimental and control group had two lessons per week of around one hour each, with one of these lessons being supplemented by multimedia use for the experimental groups. Teachers in each school were provided with an induction session on the operation and use of the multimedia package to be used. Technical staff in each school worked with the researcher to set up and test the software so that it was available for students on the school computer network prior to use with theexperimental group classes and to ensure that access was provided only to those students (through controlling student login accounts) in order to reduce the possibility of students in the non-multimedia groups being able to gain access to the multimedia resource.

The learning resources used by the experimental and control groups were equivalent in terms of the information being conveyed, being closely related to the text and the subject syllabus specifications for the public (GCSE) examination that students were preparing for. The existing paper-based learning resources of teachers included worksheets, summaries of various elements of *Macbeth* (biographies of characters, timelines of events, etc.), drawings, illustrations, sample questions, and worked examples.

The experimental groups had one of their two lessons per week replaced by a laboratory session using the multimedia software, under the direction of their usual teacher. The control group had both of their weekly lessons with their usual teacher but had no contact with the multimedia resource used by the experimental group, although no measures were taken to prevent the two groups discussing their classes with each other, as this was impractical and could have risked inadvertently invoking the Hawthorne Effect. For each teacher, the control group(s) followed the same sequence of content study in their classes as those in the group(s) which were using multimedia.

**The content of the English Literature software**

A networkable multimedia application was provided for each school to support individual and group learning about Shakespeare's play *Macbeth*. The application was from a series of computer programmes created to support the study of a number of individual English Literature texts. Each of these applications contained text, sound, graphics, animation and interactive context-sensitive elements and graphically rich interface designs which provided the student with access to relevant information, support, and learning resources which they could choose to operate in either a 'teaching' or 'testing' mode as desired.

The content of the multimedia resources covered: the text and its main literary sources; interpretations and explanations of dramatic structure and style, characters, themes, images and language; an interactive presentation of the text itself (if free ofcopyright) together with features allowing for its study, interrogation, and understanding; relevant cultural, scientific, and philosophical background to the text and for the historical period; chronological sequences and explanations of the plot and main subplots; explanatory photographs, charts, diagrams, and maps for relevant locations and events; and biographical information about the author as related to the text.

For each area of content, the resources provided a range of questions in different interactive formats (closed, open ended, cloze, true-false, multiple choice, odd-one-out, etc.) that incorporated feedback to correct, partially correct, and incorrect answers. Feedback to question answers was designed to promote further learning by offering affirmation, reinforcement for correct answers, further learning, suggestions, prompts, and hints or a factual answer, and it often also directed the learner to another area of the resource for additional help or information and to specific locations within the physical text being studied.

The Shakespeare title was selected above others after consultation with faculty because they felt that *Macbeth* provided a particularly suitable challenge for many learners in terms of its mature thematic content, its subtlety of characterisation and imagery, its cultural specificity (historical setting), and its complex language containing multiple layers of meaning. Faculty agreed that *Macbeth* was typical of subject matter that imposed on learners unavoidably high intrinsic cognitive load and was commonly seen by learners as "difficult".

The multimedia application was custom-built to the common design template that was used for the series from which the resource was taken. Packages in the series were available for a number of the most popular texts prescribed for use in many examinations at Key Stage 4 (K11-K13): George Orwell's *Animal Farm*; Charles Dickens' *Great Expectations*; William Golding's *Lord of the Flies*; Shakespeare's plays *Macbeth* and *Romeo and Juliet*; John Steinbeck's *Of Mice and Men*; Mildred D. Taylor's *Roll of Thunder, Hear My Cry*; and Harper Lee's *To Kill A Mockingbird*.

Each multimedia package contained a range of resources for exploring and understanding content and for the teaching and testing of many elements within the literary work, including its story (or plot), language, characters, themes and imagery, its dramatic structure, and its cultural context. These computer resources covered similar areas to the non-multimedia resources used by teachers in the four schools, but, additionally, included interactive multimedia features such as: spoken commentaries accompanying illustrations; interactive maps and timelines about the story; interactive audio translations of archaic or specialised technical language; hyperlinks between the text and explanations of relevant cultural

background or links to similar themes or images; and discussions or definitions of relevant dramatic features or ideas, as well as pre-programmed feedback responses (such as context sensitive responses to questions answered) in a graphic user interface featuring pictorial menus and icons designed around screens with interconnected elements.

**Instrumentation**

Previous studies have used subjective mental effort ratings to measure overall cognitive load (see Paas, 1992 Paas, F. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84: 429–434. ; Van Gog & Paas, 2008 Van Gog, T. and Paas, F. 2008. Instructional efficiency: Revising the original construct in educational research. *Educational Psychologist*, 43: 16–26. ) and found that learners who experienced the same overall cognitive load achieved different learning outcomes (Tabbers et al., 2000 Tabbers, H.K. and Martens Van Merriënboer, R.L. J.J.G. *Multimedia instructions and cognitive load theory: Split-attention and modality effects*. Paper presented at the National Convention of the association for Educational Communications and Technology. February. ). This may be due to increases in extraneous load being accompanied by decreases in germane load with some learning formats and vice versa with other learning formats, or to the prior learning of students, or their intrinsic abilities or interest in the topic.

To test whether differences in learning outcome are caused by germane load attributable to the format of instructional materials, a closer measurement of cognitive load is required. Using the text of *Macbeth* as the target area of study, differences in the extraneous cognitive load imposed by two different instructional designs were explored with one set of groups using multimedia (experimental) and another (control) using teachers' established, routine approaches to teaching the same subject matter. It was expected that the germane load imposed on students by these different approaches could therefore vary. A measure of prior knowledge was used to allow for individual differences in intrinsic cognitive load attributable to learner expertise.

Subjects were asked to record how difficult it was for them to learn using different resources, by rating the ease of learning (i.e., mental effort) they associated with different instructional materials for the plot, characters, and themes and imagery on a 5-point Likert-type scale from 1 (*extremely easy*) to 5 (*extremely difficult*). For practice, subjects rated 23 exemplar teaching resource materials for a non-Shakespeare title, representative of the repertoire of resources used by teachers in the four schools, for the degree to which subjects found them easy to learn from; they also similarly rated 32 elements from a non-Shakespeare multimedia title from the same series as the *Macbeth* resource.

Teachers within the four schools checked their existing and the multimedia resource content against a collectively constructed measure of knowledge and understanding which was applied pre- and post-intervention to ensure that measures of relative change in scores could be meaningfully used for both experimental and control groups and that the test included measures of knowledge and understanding that could be gained equally from all resources.

The test was a 30-item paper-based assessment that was completed individually by participants in the normal class session immediately preceding the start of the intervention and before the study of the selected text began. There were 10 questions about key characters in *Macbeth*, 10 about important themes and images found in the play, and 10 about its

structure and cultural context. Each section of the test included questions in open-ended, multiple-choice, and short-essay formats. The post-test used this same instrument and was completed under the same conditions 10or 11 weeks later (variations being due to differing lengths of half terms between schools).

In order to minimise the possibility of a Hawthorne Effect, where an increase in "productivity" (scores on the knowledge and understanding test) could be produced by the psychological stimulus of being singled out and made to feel important (Franke & Kaul, 1978 Franke, R.H. and Kaul, J.D. 1978. The Hawthorne experiments: First statistical interpretation. *American Sociological Review*, 43: 623–643. ), students using the multimedia application were told that this was simply one of several resources used for studying the text, that they may or may not find it useful but that, by itself, it was not expected to make any real difference to their learning, as this would be mainly the result of how hard they worked in class and on their homework study tasks. Students were not informed of their results from the pre-test, or that the test would be re-used at a later date.

## Results

### Overall results and results by school

Table 2 presents the overall results of the pre-test and post-test, for all the schools combined and for each school.

Table 2 indicates that the mean difference between the scores of the experimental groups (multimedia) on the pre-test and the post-test were statistically significant ($\rho = .000$). Similarly, the mean difference between the scores of the control groups (non-multimedia) on the pre-test and the post-test were statistically significant ($\rho = .000$). In other words, both the experimental and the control groups recorded statistically significant increases in their scores between the pre-test and the post-test.

Table 2 indicates that the experimental groups recorded a higher *difference* than the control groups in the gains made in their scores between the pre-test and the post-test. Whilst the experimental groups recorded a mean gain of 21.48 points, the control groups recorded a mean gain of 19.07 points.

The post-test difference, after removing the natural maturation, is calculated thus: {Experimental group post-test minus pre-test} minus {Control group post-test minus pre-test} = {27.04 − 5.56} minus {24.63 − 5.56} = 21.48 − 19.07 = 2.41. In other words, the difference in the mean gains between the two groups was 2.41, that is, an 8.03% difference.

For the pre-test, no statistically significant difference was found between the overall means of the experimental and control groups ($\rho > .05$), that is, they were matched at the start of the experiment), whereas for the post-test the difference between the overall means of the experimental and control groups was highly statistically significant ($\rho = .000$).

A measure of effect size (using Cohen's *d*) of the difference between the experimental and control group on the post-test yielded an effect size of $d = .896$, which is a moderate to strong effect. Using another calculation of effect size – partial eta squared – this yielded an effect size of $\eta_p^2 = .185$, which is a moderate effect. For the mean difference between pre-test and post-test intervention scores (a measure of gain), the partial eta squared yielded an effect size

of $\eta_p^2 = .989$, which is very strong (Cohen, 1988 Cohen, J. 1988. *Statistical power analysis for the behavioural sciences*, (2nd ed., New York, NY: Academic Press. ).

In reviewing the measures of statistical significance and effect size, then, the difference in scores for knowledge and understanding of Shakespeare's *Macbeth* of the groups using the multimedia resource, as compared to the scores of students who did not use the resource, was highly statistically significant and produced a moderate to strong or very strong effect, depending on the measure used. However, the size of an effect, which is often more informative and for practical purposes more relevant than a statistical correlation, also needs to be distinguished from its *importance*. It is necessary to consider what these effect sizes mean in terms of the units of measure of the original variable which, in this case, is the improvement in knowledge and understanding that relates to the subject syllabus for an external examination.

Grades for GCSE examinations in the UK are allocated using bands of marks with different ranges delimiting the grades at Higher or Foundation levels and the boundaries for ranges vary slightly between Examination Boards; those referred to here reflect the syllabus in use in the four schools. Table 3 indicates the grade boundaries used in the GCSE examinations.

The results reported here therefore indicate that teachers, by making use of the multimedia resources discussed, could have improved their students' average performance in English Literature by a margin in excess of one GCSE examination grade.

Though these gross differences were found in combining the results from the four experimental groups and in combining the results from the four control groups, nevertheless differences were found between each of the experimental groups in eachof the four schools and between the control groups in each of the four schools. Table 2 indicates where the difference lay between the four schools.

To ascertain whether there were any statistically significant differences between the schools at the pre-test stage, Analysis of Variance (ANOVA) and the post-hoc Tukey test were conducted on the four experimental groups and on the four control groups. ANOVA found that there was a statistically significant difference between the four schools at the pre-intervention (pre-test) stage for both the experimental and control groups ($F = 3.950$, $\rho = .009$). The Tukey test found that the means for the four schools were statistically significantly different at the pre-intervention stage with the mean for School 1 being relatively low (5.13), which was some distance away from the next lowest mean of 5.53 (School 3).

At the post-test stage, ANOVA found that the means for the four Schools in the experimental (multimedia) group were not statistically significantly different ($\rho = 0.966$), whereas in the control (non-multimedia) group the means were statistically significantly different ($\rho = .003$), with the mean for the control group in School 1 (23.95) being some distance away from the next lowest mean (School 4: 24.83) and with the mean for the control group in School 3 being some distance away from the next highest mean (School 2: 24.89). Though one can see improvements in all the schools and for both the experimental and control groups, Table 2 suggests that School 4 made the greatest difference to both Groups 1 and 2, though these were only a little higher than those in the other three schools.

A closer inspection of the data revealed that statistically significant differences between the mean scores for mixed-sex schools at both pre- and post-intervention stages are largely the product of performance by sex, although there was no consistent direct relationship between sex, an individual school, and the mean gains for multimedia or non-multimedia groups.

Explaining the variability between students within the four schools is of interest, especially for teachers seeking to understand why measures of overall gain varied widely between and within schools. For example, the pre-test mean score for knowledge and understanding was lower for students in School 1 (5.13) – both forthe multimedia (experimental) and non-multimedia (control) groups – than for students in any of the other schools, but the use of multimedia in this school produced the largest absolute increase in average scores over the students in the non-multimedia group in all schools (3.07). The use of multimedia in the single-sex school (School 3) produced the smallest absolute increase in average scores over the non-multimedia group in any of the four schools (1.88), despite this school having one of the highest overall pre-intervention scores for knowledge and understanding of *Macbeth* for both multimedia and non-multimedia groups (5.83).

These data suggest that the use of the multimedia resource in the four schools improved scores of knowledge and understanding in English Literature substantially beyond those achieved without the resource, but there was no consistent linear relationship between the pre-intervention scores and post-intervention scores (relative gain) of students that used the multimedia resource.

**Results by sex**

GCSE examination pass rates have continued to rise in the UK over the last 23years, but there remains a persistent and, in some subjects (especially English), agrowing difference in performance between males and females, although at "ALevel" this gap is now closing. This phenomenon continues to attract comment in both the media (BBC, 2011a BBC. 2011a. *GCSE results: Gender gap widens in record-breaking year* Retrieved from http://www.bbc.co.uk/news/education-14661746, 2011b BBC. 2011b. *Why boys trail further behind girls at GCSE top grades* Retrieved from http://www.bbc.co.uk/news/education-14664916; Guardian, 2011 Guardian. 2011. *GCSE results 2011: exam breakdown by subject, school and gender* Retrieved from http://www.guardian.co.uk/news/datablog/2011/aug/25/gcse-results-2011-exam-breakdown; Mail Online, 2011 Mail Online. 2011. *March of the girls! GCSE day sees record results ... but boys fall futher behind as gender gap hits record level* Retrieved from http://www.dailymail.co.uk/news/article-2029847/GCSE-results-2011-Record-results–boys-fall-girls.html) and in government reports (DfCSF 2007, 2009). The data were therefore processed and analysed by the sex of the students (Tables 4 and 5). The knowledge and understanding assessment was a 30-item test, and the average gains in scores in the four schools ranged from 6.27% (School 3) to 10.23% (School 1), with the highest score being for females in School 1 (11.47%) and the lowest being for males in School 2 (5.47%).

The overall pre- and post-test scores for males and females in the experimental and control groups, for schools, and for males and females within each school were examined for differences between the pre-test and post-test, using the *t* test for independent and related samples as appropriate.

On the English Literature knowledge and understanding variable, the results are shown in Table 6. For males the mean score (from a maximum score of 30) for the pre-test was 5.44 (*SD* = 1.864); for the post-test, it was 25.51 (*SD* = 2.924), and themean percentage improvement between pre-and post-test scores was 66.88 (*SD* = 8.565). For females, the mean score on the pre-test was 5.62 (*SD* = 1.927), and for the post-test it was 26.02 (*SD* = 2.720), with a mean percentage gain of 68.00 (*SD* = 7.985).

Tables 4–6 indicate that, for the pre-test, the results of the males and females did not differ statistically significantly ($\rho > .05$). For the post-test, the results of the males and females differed statistically significantly ($\rho = .015$). There was a highly statistically significant difference between males and females in the sample for the overall pre-test to post-test scores and for the relative measure of learning gain in knowledge and understanding, with females achieving larger gains (8.33%) relative to males (7.47%). A similar pattern obtained in each of the three mixed-sex schools, where females using the multimedia resource gained larger improvement than males using the multimedia resource.

In School 2, females scored higher than males at both pre-test and post-intervention stages in both the control and experimental groups. In School 1, females scored higher than males in the experimental group, but in the control group males scored higher than females, although the mean gain of the experimental group as awhole (22.22) was still greater than that of the control group as a whole (18.78), and a similar situation existed in School 4. In control groups, knowledge and understanding scores improved more for males than females in School 1, but these gains were reversed and equalled or exceeded in size by females in all the other schools.

Females in the single-sex school (School 3) achieved the smallest mean gain for females in the experimental group (21.21) out of all four schools, whilst females in School 1 achieved the largest gain (22.22). School 4 achieved the largest mean gain for females in all schools in the control group (19.50), and School 1 achieved the smallest (18.78). Males in School 2 achieved the smallest mean gain for males in the experimental group in the three mixed-sex schools (20.65), whilst males in School 1 achieved the largest mean gain for males in all schools in the control group (21.53). Overall, School 1 produced the largest mean gain for all students using multimedia (21.89) (i.e., experimental group), and the single-sex School 3 produced the largest mean gain for all students using non-multimedia (19.33) (i.e., the control group), closely followed by School 4 (19.30).

**Interaction effects: sex, school, and group**

A two-way between-group analysis of variance was conducted to discover the interaction effects of sex, school, and group (experimental or control) on performance in the examination of GCSE English Literature competence. There was a statistically significant main effect for experimental/control group ($F = 200.816$, $\rho < .001$, $\eta^2 = .206$). There was also a statistically significant effect for sex ($F = 6.076$, $\rho = .014$) and for the interaction effect for school * group ($F = 4.350$, $\rho = .005$), but the effect sizes were small (sex $\eta_p^2 = .008$; school * group $\eta_p^2 = .017$). The interaction effects and effect sizes were also not statistically significant and weresmall as between school * sex ($F = 0.211$, $\rho < .810$, $\eta_p^2 = .001$); group * sex ($F = 146.156$, $\rho < .089$, $\eta_p^2 = .004$); and school * group * sex ($F = 8.962$, $\rho < .837$, $\eta_p^2 = .000$) (Cohen, 1988 Cohen, J. 1988. *Statistical power analysis for*

*the behavioural sciences*, (2nd ed., New York, NY: Academic Press. ). Table 7 indicates that the use or non-use of multimedia was the most important and substantial influence on improvements in learning the course content studied in each of the four schools ($\eta_p^2 = .206$). Put simply, the multimedia intervention exerted a stronger effect on the post-test scores than either the sex of the student or the school which s/he attended.

**Cognitive load and resources**

The teaching and learning resources were varied, and cognitive load theory (CLT) suggests that different kinds of resource would make different demands on students. It was important to explore the kinds of demands made on students by the different resources, including those in the multimedia package. To ascertain the nature of the cognitive load and demands made on students using the multimedia resources (the experimental group), factor analysis was conducted on the data concerning the different kinds of resources that they used.

Further, CLT offers a possible explanation for the lack of a direct consistent relationship found between pre-test scores and post-test scores by arguing that the critical factor in facilitating successful learning is the degree to which cognitive loadis optimised and that, as discussed in the earlier part of this article, this is the product of: instructional design (extraneous cognitive load); the effort expended in constructing, processing, and automating schemas (germane cognitive load); and prior learning and intrinsic subject/topic difficulty (intrinsic cognitive load).

Given the parallel instructional resources used by students in the experimental (multimedia) groups in the study, it was deemed reasonable to expect that *extraneous* cognitive load (load associated with the design of instructional materials and the way these present information to the learner) was held constant for these students. *Intrinsic* cognitive load is directly attributable to the inherent complexity or difficulty of the material to be learned and is composed of the combination of the learner's prior knowledge and the complexity of the learning material. For the purposes of the present study, one may assume that the part of intrinsic cognitive load attributable to the complexity of the learning material (*Macbeth*) was also constant for these students, as they were all studying the content of the same literary work. The remaining component of intrinsic cognitive load (prior learning) remains as the important variable; this was measured by reports of individual ratings of mental effort when learning with different resources (which has been shown to be an effective proxy (Kalyuga et al., 1998 Kalyuga, S., Chandler, P. and Sweller, J. 1998. Levels of expertise and instructional design. *Human Factors*, 40: 1–17. )).

Following an introductory 2-hr familiarisation period provided by the researcher and the class teacher using an alternative title from the multimedia series of packages, individual subjective ratings of mental effort were obtained for learning with a range of its relevant features and also for learning with a range of exemplar classroom resources commonly used by the teachers from the experimental (multimedia) groups in the four schools, using a paper-based questionnaire and accompanying illustrated workbook. Ratings were used as proxy measures of the intrinsic cognitive load attributable to the learning materials (equivalent to

learner expertise) and compared with the learning gains made by individuals within the experimental (multimedia) groups in the four schools.

Twenty-nine different examples of resource were assessed, chosen to be representative of those used by the teachers in the four schools and contained within the multimedia software: (a) text only; (b) text accompanied by a number of graphical features that were integrated with the text; (c) text accompanied by unintegrated graphical features that seemed likely to invoke a split-attention effect; and (d) integrated multimedia elements from the multimedia resource.

Exploratory principal component factor analysis was used to investigate the underlying constructs within the data from the students concerning these resources. It was expected that the mean scores for different individual resources would not all intercorrelate; hence, Varimax rotation was applied to the data, and the rotated component matrix maximised the differentiation of the original variables (different resources) by extracted factor. From the data, four clear components were extracted with Eigenvalues greater than 1.00, and these accounted for 64.489% of the total variance explained.

Small coefficients were initially suppressed in the rotated component matrix at the .50 level, but one item (non-multimedia non-interactive maps – P5 NM Maps) did not initially appear to load onto any component at this value, but increasing this to an absolute value of below .10 produced an output for P5 NM Maps loading most strongly onto Component 1, and this additional value was therefore added to the output (Table 8).

The four components were entitled thus:

- Factor 1 (Component 1): Non-multimedia text resources that were accompanied by illustration elements that were integrated with the text (22.055% of total variance explained);
- Factor 2 (Component 2): Non-multimedia text resources unaccompanied by other features (16.759% of total variance explained);
- Factor 3 (Component 3): Non-multimedia text resources that were accompanied by illustration elements that were not integrated with the text and which were likely to produce a split-attention effect (15.699% of total variance explained);
- Factor 4 (Component 4): Multimedia resources with integrated elements such as animation and/or speech (9.976% of total variance explained).

These four factors were scored by students for the degree of difficulty they experienced when using them for learning for each of the factors (on a 5-point scale), by aggregating the means for each of the elements in each factor. The results are reported in Table 9. Researchers in the field of CLT have demonstrated that suchscores are reliable proxies for learner expertise and thus for the level of germane load imposed on the learner by instructional materials (Cierniak et al., 2009 Cierniak, G., Scheiter, K. and Gerjets, P. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?. *Computers in Human Behavior*, 25: 315–324. ; Kalugaet al., 2003, 1998; Salomon, 1984 Salomon, G. 1984. Television is “easy” and print is “tough”: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76: 647–658. ). As extreme levels of cognitive load are detrimental to learning (Van Merriënboer & Sweller, 2005 Van Merriënboer, J.J.G. and Sweller, J. 2005. Cognitive load theory and complex learning: Recent developments and

future directions. *Educational Psychology Review*, 17: 147–177. ; Young & Stanton, 2002 Young, M.S. and Stanton, N.A. 2002. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, 44: 365–375. ), one would expect that variations in scores for these resources would be closely associated with (i.e., would predict) measures of learning gains when controlled for resource type (component). CLT would predict that students who reported that they found a particular resource very difficult to learn with would gain less knowledge and understanding from using it than students who reported that they found the same resource very easy to learn with.

Table 9 indicates that both the experimental and control groups found that Factor 4 – multimedia resources with integrated elements such as animation and/or speech – provided the greatest ease and the lowest degree of difficulty, whereas Factor 2 – non-multimedia text resources unaccompanied by other features – provided the greatest degree of difficulty. Regardless of whether a student was in the experimental or control group, the same progression of ease was found, from hardest to easiest: Non-multimedia text resources unaccompanied by other features was the most difficult, followed, as second most difficult, by non-multimedia text resources that were accompanied by illustration elements that were not integrated with the text and which were likely to produce a split-attention effect, followed as third most difficult (second easiest) by non-multimedia text resources that were accompanied byillustration elements that were integrated with the text, and the easiest was multimedia resources with integrated elements such as animation and/or speech. Clearly, the more integrated and animated were the resources, the greater was the easing of cognitive load, and the use of multimedia (Factor 4) provided the greatest easing of cognitive load (cf. Mayer & Moreno, 2002 Mayer, R.E. and Moreno, R. 2002. Aids to computer-based multimedia learning. *Learning and Instruction*, 12: 107–119. ). When ANOVA and the post-hoc Tukey test were conducted on the four means of the experimental group and thefour means of the control group, the mean for Factor 4 was statistically significantly different from the means of the other three factors ($\rho = .008$) for both the experimental and control groups, that is, using the multimedia software made a statistically significant difference to the cognitive load, easing it, for both the experimental and control groups. This was particularly the case for the experimental group, where the mean for Factor 4 (3.64) was substantially lower than that of the control group (3.71) and showed the greatest difference between the experimental and control groups of all the four factors (see Table 9); for the experimental group, the mean for Factor 4 (3.64) was .17 distant from the second lowest score (3.81), and for the control group it was .15 distant from the second lowest score (3.86). Both of these distances are in stark contrast to the distances between the second lowest and the highest scores for each group: .11 for the experimental group and .06 for the control group.

**Discussion**

The findings from the research are very clear: In the English Literature study, cognitive load was eased by the use of multimedia, and the greatest easing was where multimedia were integrated with animation, audio, explanation, and background analysis. Further, the study indicates that CLT can offer both a useful explanation for the findings, and it can assist teachers of English Literature in planning theirteaching, particularly in terms of the resources that they use. The study has indicated that intrinsic cognitive load in English Literature teaching and learning canbe rendered more manageable by the use of integrated and

advanced multimedia that move beyond text to animation, background materials, and voice-over commentaries.

CLT argues that *germane* cognitive load (the load directed towards constructing, processing, and automating schemas) can be manipulated and optimised by good instructional design in ways that help learning by directing attention more towards relevant learning processes. The findings of the present research show that this outcome was achieved more by the multimedia resources used than by the other resources used and that this benefit was progressive across the different media in line with the expectations of CLT. This finding supports those found in other studies mentioned earlier (Cierniak et al., 2009 Cierniak, G., Scheiter, K. and Gerjets, P. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?. *Computers in Human Behavior*, 25: 315–324. ; Kalyuga et al., 2003 Kalyuga, S., Ayres, P., Chandler, P. and Sweller, J. 2003. The expertise reversal effect. *Educational Psychologist*, 38: 23–32. , 1998 Kalyuga, S., Chandler, P. and Sweller, J. 1998. Levels of expertise and instructional design. *Human Factors*, 40: 1–17. ; Mayer 2008 Mayer, R.E. 2008. Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63: 760–769. , 2009 Mayer, R.E. 2009. *Multimedia learning*, Cambridge, , UK: Cambridge University Press. ; Mayer & Moreno, 2002 Mayer, R.E. and Moreno, R. 2002. Aids to computer-based multimedia learning. *Learning and Instruction*, 12: 107–119. ; Schnotz & Kürschner, 2007 Schnotz, W. and Kürschner, C. 2007. A reconsideration of cognitive load theory. *Educational Psychology Review*, 19: 469–508. ). In comparison tothe other (non-multimedia) resources used, including those customarily employed by the teachers in the four schools, the multimedia also produced lower *extraneous* cognitive load on working memory which, if it is too high, is unhelpful for learning and is created as a result of unnecessary processing caused by instructional design. The findings of the present research also show that, because ofthe above outcomes, the multimedia resource was helpful in moderating the effects of the *intrinsic* cognitive load attributable to the inherent complexity and difficulty of the material to be learned (*Macbeth*) by offering learners who had relatively little prior knowledge of this literary work the integrated resources that CLT predicts will be most helpful in facilitating the creation of the schema required for learning.

## Conclusions and implications

The study found appreciable differences in cognitive load between the experimental and control groups and that the use of multimedia was strongly associated with increased performance in important areas of learning most relevant for success in external examinations at GCSE. The effect sizes and statistically significant differences found between the control groups and the experimental groups here were notable, being equivalent to around one grade difference in GCSE results. Further, statistically significant differences were found within and between the experimental and control groups by sex, with females typically outperforming males. However, the multimedia intervention exerted a stronger effect on the post-test scores than either the sex of the student or the school which s/he attended, and the multimedia showed the greatest difference/distance between the mean scores for other kinds of resource for all the other factors. In other words, the use of multimedia was associated with the strongest differences of all those found.

The use of the multimedia resource in the four schools improved scores in the field of knowledge and understanding in English Literature substantially beyond those achieved without the resource. Using the multimedia software made a statistically significant

difference to the cognitive load experienced by students, easing it for both the experimental and control groups when compared to the cognitive load associated with other resources.

That said, caution has to be exercised in interpreting the results here. For example, there were between-group differences in the four schools in terms of the control groups and in terms of the experimental groups, and this suggests that the influence of cognitive load resides not only in external matters, such as the teaching and learning materials used (though clearly they make a significant difference) or the types of materials within multimedia (which the tables will show), but that it also resides within the learners. The ease of learning with any given educational resource may be in part due to the resource itself, or to the prior learning of the individual, but also to a host of other subjective, personal, or biographical factors such as student motivation, student teacher interaction, or student'student interaction, and these other factors cannot be ruled out as having no bearing on the results found, and we cannot be certain that cognitive load is a feature of only "cognitive" processes. Further research needs to be done on what these other factors may be and what may be their relative weightings on outcomes.

Further, though controls were exerted wherever possible between the control and experimental groups, this was a quasi-experiment and, being a field experiment, it was not therefore possible to hold some variables constant, that is, to have complete controls in place. For example, it was not always possible to match exactly the control and experimental groups (e.g., one school was single-sex), and random allocation was not possible even though several steps were taken to ensure matching between the control and experimental groups, as discussed in the article. Further research with fuller controls would be valuable here, although caution will be needed to ensure that any outcomes from strictly controlled experimental studies are generalisable.

Finally, the research operates on largely a "black box" situation here; it is concerned largely with input and output measures. This does not tell us about the effects of process variables and of variables within the learners (and, as discussed above, subjective factors are relevant in analyses of cognitive load), and these may be important. Hence, the present article not only presents findings but raises questions for further research, especially concerning those process variables and learner-related variables that could be exerting an influence on the situation, which could be researched, for example, through observational studies.

However, despite these limitations, the findings here suggest that educationists who are keen to discover how to make cognitive load manageable, particularly for the teaching and learning of inherently complex matters, will find the use of active and interactive multimedia techniques valuable in easing that degree of cognitive load which is counter-productive to effective and efficient learning. This article has suggested that cognitive load theory can make a significant contribution to understanding the limiting constraints on students' learning brought about by instructional design and its associated cognitive load and overload, and that cognitive load theory can suggest effective ways of how to reduce and ease these.