Authors

Philip C. Abrami, Robert M. Bernard, C. Anne Wade, Eugene Borokhovski, Rana Tamim, Michael Surkes, and Dai Zhang

Centre for the Study of Learning and Performance, Concordia University, Montreal, Quebec

Abrami et al. (this issue) provide a review of e-learning in Canada from 2000 onwards by synthesizing information drawn from multiple sources, not only primary research. In total, there were 726 documents included in our review: 235 views expressed in the public printed media (an expression of general public opinion); 131views from trade/practitioner perspectives; 88 views of policy-makers contained in public policy documents; 120 sources of evidence contained in reviews of research; and 152 sources of evidence contained in primary empirical studies. As far as we know, this is the first review of its kind to be as inclusive of sources. This is both our review's greatest strength, allowing us to determine whether consensus existed among myriad sources, and potentially its greatest weakness, as both the time for such a comprehensive review (our contract with CCL specified the review needed to be completed in 90 days) and the cost meant we sacrificed a degree of depth for an increase in breadth.

We reached a number of conclusions in our review, a few of which bear repeating:

- Conclusions from Canadian primary research, international literature reviews, policy documents, media reports, and practitioner publications are mostly favourable towards the use and impact of e-learning (i.e., student achievement, motivation, and other outcomes) in Canada.
- In Canada there is a lack of evidence in some theme areas, notably early childhood learning, and a lack of experimental and quasi-experimental evidence that would allow unambiguous causal conclusions to be drawn about effectiveness.
- The quality and scope of the research evidence does not match the time, cost and resources that have been and will be dedicated to the development and implementation of e-learning.
- There is a need for programs of development for new initiatives that have high-quality research and evaluation programs or components built-in as a forethought, not an afterthought.
- And we expressed important limitations, some of which also bear repeating:
- There is an unanswered question concerning the emphasis which is placed on deployment, attendant costs, and what one might take away, or not add, given the expense of an e-learning delivered curriculum.
- There are promising areas of new development focusing on specific applications of technology such as: learning objects and repositories; standardization of metadata; electronic portfolios; and broadband enabled lifelong learning projects.
- We emphasized Canadian primary research to determine the nature and extent of evidence in our country. Our ability to compare Canadian evidence with research from other countries was limited and indirect. We included only literature reviews of non-Canadian primary research.
- Because of the scope of this undertaking, its novelty, and the time constraints under which we operated, we have been able to provide only a rough portrait of the evidence and opinions.

- While we incorporated a large number of explanatory variables or study features, we are certain that a finer analysis of the literature would yield far more.
- We did not examine the evidence from a theoretical perspective, in part because of the time limits, but primarily because there is little in the way of theory-testing research on e-learning that can be synthesized.
- There were methodological challenges and shortcomings to this review and to our use of an argument catalogue to synthesize views on e-learning.
- Our analyses of evidence, including the primary evidence, are based primarily on frequency analyses or vote counts of impacts without regard to the methodological quality of the evidence.

After we completed our review for the Canadian Council on Learning, we contacted Michele Jacobsen, the Editor of the *Canadian Journal of Learning and Technology*, not only to learn whether the Journal was interested in publishing our work but also to see whether the paper merited enough attention to request comments from Canadian experts to which we could reply. We appreciate the willingness of the journal to publish our work and we are especially grateful to Terry Anderson, Margaret Haughey, Heather Kanuka and Richard Schwier for examining our work in such detail. While we do not agree on several key points, we believe this printed dialogue does much to improve our thinking and to advance understanding amongst us all.

In our rejoinder to the commentaries, we explore several issues: a) the nature and importance of systematic reviews; b) answering questions about what works and why; and c) what is e-learning and what are its impacts. In doing so, we do not address every point raised in the commentaries, some of which we agree with, some of which are also covered in our review, and some of which we disagree with.

The nature and importance of systematic reviews

In her commentary, Kanuka (this issue) notes that:

> Missing in the review on achievement are research findings which have revealed students infrequently engage in the communicative processes that comprise critical discourse—an essential component of achievement as it relates to higher levels of learning (see for examples: Aviv, Zippy, Ravid & Geva, 2003; Bonk & Cunningham, 1998; Bullen, 1999; Davis & Rouzie, 2002; De Laat, 2001; Garrison, Anderson & Archer, 2001; Gunawardena, Carabajal & Lowe, 2001; Gunawardena, Lowe & Anderson, 1997; Jeong, 2004; Kanuka, 2005; Kanuka & Anderson, 1998; Lopez-Islas, 2001; McKlin, Harmon, Evans & Jones, 2002; McLaughlin & Luca, 2000; Meyer, 2003; Nussbaum, Hartley, Sinatra, Reynolds & Bendixen, 2002; Pawan, Paulus, Yalcin & Chang, 2003; Pena-Shaff, 2005; Pena-Shaff, Martin, & Gay, 2001; Pena-Shaff & Nicholls, 2004; Rourke, 2005; Rovai & Barnum, 2003; Thomas, 2002; Vaughan & Garrison, 2005; Veerman, Andriessen, & Kanselaar, 2000; Wilson, et al., 2003; Yakimovicz & Murphy, 1995). Research conducted by Angeli, Valanides, and Bonk (2003) is representative of many of these studies' conclusions: "students primarily share personal experiences amongst themselves, and their responses appeared to be subjective and naïve at times. Students' discourse was also extremely conversational and opinionated and showed little evidence of critical thinking" (p. 40). It has been difficult for most of us concerned with e-learning in higher education to ignore these disappointing results—and, yet, these findings have not been reflected in the team's review of the literature. (p. 88).

Kanuka's conclusion may or not be correct, but it provides us an opportunity to examine critically how she reached it. We do so not to address a criticism at our colleague but instead to underscore the importance of describing and using systematic procedures when reviewing research.

Is the evidence complete? The evidence and the citations given in support of Kanuka's conclusion represent a limited review of the literature on the relationship between

students' online communication and critical discourse, posited as an aspect of higher-order learning and critical thinking. Because there are many studies cited which appear to reach this result, Kanuka concludes that these results are firmly held, implying that there are few, if any, courses, students, and contexts where the findings would be otherwise. This is a form of review by vote counting that we will comment on later in our rejoinder. At first glance, Kanuka seems to ignore the research on Scardemalia and Bereiter's Knowledge Forum (1996) and Feenberg's Textweaver (n.d.), which we mentioned at the end of our review, among others. We wonder whether other studies showing positive results have also been excluded. For example, very recently a call for submissions to *Contemporary Educational Psychology* on Collaborative Discourse, Argumentation and Learning states: "There is a small but growing body of evidence that collaborative student discourse (i.e., reflective discussions among students about academic content) can promote deep and meaningful learning and enhance students' reasoning skills" (2006). In fact, Kanuuka (2005) reports that her own action-based research showed "evidence of higher levels of learning resulted from [synchronous online discussion]" (Brainstorming section, ¶ 5).

The importance of systematicity. A systematic review attempts to be objective, repeatable, and transparent by avoiding the subjective and idiosyncratic weaving together of one's impressions of evidence. Our worry about many narrative reviews which do not follow systematic procedures is the increased probability of divergent interpretations of different collections of evidence leading to confusion and frustration among researchers, policy-makers and practitioners.

So in the case of the systematic reviews of the primary research that we have conducted and reported on (e.g., Bernard et al. 2004), and by extension used in the Argument Catalogue, we are careful to lay out the steps we followed in the conduct of the review and to explain carefully how we proceeded at each step. These review steps include: a) identifying and explaining a core question; b) systematically searching the literature; c) articulating inclusion and exclusion criteria: d) extracting key indicators of effect or outcome; e) coding study features; and f) summarizing key findings and exploring variability among the results. In a meta-analysis, step d) involves extracting effect sizes (i.e., quantitative indicators of standardized mean difference between a treatment and a control group), but they may take other forms in other kinds of reviews as they did in our use of the Argument Catalogue. Therefore, we contend that, regardless of whether a quantitative review, a qualitative (narrative) review, or an Argument Catalogue is being conducted, these steps are important to follow because systematicity is an important aspect of objectivity, which also maximizes inclusiveness and fairness and minimizes bias.

In the case of Kanuka's selection of references, we see that the precise meaning of her core question is unclear, there is no evidence of any attempt to systematically search the literature, and there is nothing to describe why studies are included or what studies are excluded from consideration. In fact, there is a mismatch between the studies in Kanuka's citations and the inclusion criteria we employed. Some of her studies were conducted prior to 2000, some are not Canadian in origin, and so on. It is not surprising, then, that her

conclusions on this point are at odds with ours. Table 1 summarizes whether each of Kanuka's 28 studies should have been included in our review and the reasons why many of them were excluded. It appears that, in fact, we did not miss a large collection of studies using our stated inclusion criteria. Had we not explicitly described our inclusion/exclusion criteria, we could not draw this conclusion with any certainty.

Table 1. *Twenty-eight studies on online learning and critical discourse*

Citation	Comment*
Angeli, Valanides & Bonk, 2003	3
Aviv, Zippy, Ravid & Geva, 2003	3 (not in English)
Bonk & Cunningham, 1998	2
Bullen, 1999	2
Davis & Rouzie, 2002	3
De Laat, 2001	3 & 4
Garrison, Anderson & Archer, 2001	5
Gunawardena, Carabajal & Lowe, 2001	5
Gunawardena, Lowe & Anderson, 1997	2
Jeong, 2004	3
Kanuka, 2005	1
Kanuka & Anderson, 1998	2
Lopez-Islas, 2001	3
McKlin, Harmon, Evans & Jones, 2002	3
McLaughlin & Luca, 2000	3
Meyer, 2003	3
Nussbaum, Hartley, Sinatra, Reynolds & Bendixen, 2002	3
Pawan, Paulus, Yalcin & Chang, 2003	3
Pena-Shaff, 2005	3
Pena-Shaff, Martin & Gay, 2001	3
Pena-Shaff & Nicholls, 2004	3
Rourke, 2005	6
Rovai & Barnum, 2003	1
Thomas, 2002	3
Vaughan & Garrison, 2005	7
Veerman, Andriessen, & Kanselaar, 2000	3
Wilson, et al. 2003	1
Yakimovicz & Murphy, 1995	2

*Legend
1. Included in Abrami et al.
2. Study was published prior to 2000 (not in the inclusion criteria).
3. Study was conducted outside Canada (not in the inclusion criteria).

4. Study was conducted in the workplace (not in the CCL mandate).
5. Study was not indexed as Canadian and marked as a 141 (Reports-Descriptive) in ERIC (search strategy only retrieved 142 (Reports-Evaluative) and 143 (Reports -Research).
6. This dissertation should have been retrieved (search strategy did not use the term "distance learning," but instead "Distance Education" and a variety of related terms).
7. This journal article should have been retrieved, but it is not listed in ERIC. It can be found in Google Scholar but we only searched this database for reviews, not primary studies given the large number of hits that would have been retrieved on this topic.

Restricting ourselves to these 28 studies, we need to ask whether any critical discourse was evidenced in the studies of online e-learning and how variable was this evidence? Is the magnitude of the relationship between online learning and discourse uniformly zero, or always very low? Or do aspects of instructional design, course level, student characteristics and context affect the relationship? These are the sorts of questions that any review, but especially a systematic review, should strive to answer.

Answering questions about what works and why

Kanuka's (this issue) critique of our review mentioned that we did not explicitly state our "biases and undeclared assumptions," which nevertheless emerged in our treatment of the topic. In reply, we acknowledge our positivistic philosophical approach but remind readers that we specifically avoided a restrictive approach to the synthesis of evidence and views on e-learning. We did not limit ourselves to only primary quantitative studies but also included qualitative ones. More generally, the Argument Catalogue was an integration of five sources of evidence and views and not just primary research evidence. We believe the Argument Catalogue has the potential to become a new, inclusive standard for incorporating multiple perspectives on an issue. In some ways we are disappointed that we did not communicate well enough the value and importance of this approach to reviewing, where we not only included multiple sources but attempted to represent the key findings from these sources in both quantitative terms (i.e., vote counts and effect sizes) and qualitative terms (i.e., narrative descriptions).

The remainder of this section is intended to address our philosophical and methodological concerns, and to elaborate our considerations with regard to these important matters.

About positivism. From a positivist perspective, many scientific questions have at their core a concern about the (causal) relationship between two variables—often described as the relationship between an independent (predictor) variable and a dependent (outcome) variable. This is true in the natural sciences as well as the human sciences and is evident in the question that drives our review—what do we know about the outcomes or impacts of e-learning. We also believe this core concern is of interest to many policy-makers, practitioners, and the general public. This interest is not restricted to the United States or the What Works Clearinghouse but is evident in other countries (e.g., the Evidence Network and the EPPI Centre in the UK, the Nordic and Japanese Campbell Collaboration

Centres) and internationally (e.g., the Cochrane Collaboration and the Campbell Collaboration). In Canada, we have the Canadian Cochrane Collaboration and the CCL who are interested in what works and the promotion of evidence-based practice (see for example:http://search.ccl-cca.ca/CCL/Newsroom/Backgrounders/?Language=EN)

What Haughey (this issue), Kanuka (this issue) and Anderson (this issue) challenge is the epistemic tradition that underlies a positivist perspective to scientific inquiry. This has been an ongoing controversy in educational and social science research for years and unfortunately shows little sign of abating. Their challenge is especially relevant to aspects of our Argument Catalogue that resemble a quantitative review of empirical evidence; their challenge is less relevant to the narrative portions. In a recent defence of field experiments in distance education, Abrami and Bernard (2006) examined various forms of this challenge and explained the value and importance of educational experiments.

The purpose of Abrami and Bernard (2006) was to describe the range of issues involved in experimentation, to explain the importance of field experiments—what we can learn and not learn—and to discuss how the quality of such research can be improved so as to strengthen recommendations for "what works" in our field. We considered arguments for and against field experimentation, drawing heavily on a recent working paper by Cook (2004).

There is still a large contingent of researchers who maintain that the world is too complex and messy for experiments, and that notions of causality are too difficult to establish, to base all evidence for practice and policy making on "simplistic" positivistic approaches to empiricism. By contrast, there exists a strong resurgence of interest in randomized control trials (RCTs) as "the gold standard" of evidence, which should direct the course of future educational development—what works in education. While acknowledging the value of non-quantitative forms of research for evaluation, exploration, and hypothesis generation, we maintain that there are good reasons to continue with forms of field experimentation in education and when appropriate questions are being asked. Abrami and Bernard (2006) critically examine five arguments against experimentation: philosophical arguments; practical arguments; arguments about undesirable trade-offs; arguments that educational institutions will not use experimental results; and arguments that experiments are not necessary because better alternatives exist.

For example, philosophical arguments are designed to show that experiments: 1) cannot provide unbiased tests of causal hypotheses, and 2) are predicated on a descriptive theory of causation that is less useful than explanatory theories of cause.

Kuhn (1970) among others has argued that experiments are biased by the researcher's hopes, opinions, and expectations, thus undermining their neutrality and claims concerning truth. However, this criticism is equally applicable to qualitative research methods. Yet, this is no reason to dismiss experimentation out of hand. Many claims that are now known with a great degree of certainty (e.g., the effects of time on task) were initially established through experimentation. Not all evidence is subjective evidence, and not all evidence is

context-bound.

The second philosophical argument, one that Haughey (this issue) makes explicitly, maintains that experiments are predicated on an overly simplistic theory of causation, testing only the impact of a small subset of possible causes, often only a single one, rather than the full complexity of factors in a system of casual influences. For example, Haughey (this issue) wrote: "Context was so strong an intervening variable that it was impossible to parcel out the use of the new technologies, and the teacher's beliefs about, comfort with and pedagogical approaches to ICT were equally important and complex intervening variables" (p. 115).

This is one reason why multifactor experiments are designed, but there are practical limits to what can be manipulated and carefully controlled in the field. This view helps explain why correlational designs and experimental-correlational hybrids can be important adjuncts to simple field experiments. In addition, mixed-methodology research, attempts to combine the strengths of both approaches, generalizability and exploration of context, to bear on research problems. And the systematic literature review is an important adjunct to field experimentation by providing a comprehensive synthesis of the evidence and means to test for context effects among the collection of studies (see below).

Furthermore, while experiments may be limited, they are not useless. There are times when it is important to know whether a treatment or cluster of treatments "works," even if we don't know the exact cause(s). For example, experiments have helped establish the impact of the positive effects of summer school on achievement, the value of assigning and grading homework, and so on. It is, important to know whether e-learning "works" and especially to know the conditions under which e-learning is beneficial, harmful or of no particular value. Such an important question requires not only carefully controlled longitudinal investigations, but investigations brought to scale exploring the impact of context and sustainability over time and circumstance.

The importance of integrations. A systematic review of quantitative evidence (a meta-analysis) has the following advantages: a) it answers questions about effect size; b) it systematically explores the source of variability in effect size; c) it allows for control over *internal validity* by focusing on comparison studies vs. one-shot case studies; d) it maximizes *external validity* or generalizability by addressing a large collection of studies; e) it improves statistical power when a large collection of studies is analyzed; f) the effect size is weighted by sample size—large sample studies have greater weight; g) when a review is updated, it allows new studies to be added as they become available or studies to be deleted as they are judged to be anomalous; h) it allows new study features and outcomes to be added to future analyses as new directions in primary research emerge; i) it allows analysis and re-analysis of parts of the dataset for special purposes (e.g., military studies, synchronous vs. asynchronous instruction, web-based instruction); and j) it allows comment on what we know, what is new, and what we need to know (Abrami, Cohen & d'Apollonia, 1988; Bernard & Naidu, 1990). In short, meta-analysis goes far beyond what a

single study might ever hope to contribute about a phenomenon and provides a greater case for the generalizability of results across populations, materials, and methods. It also allows us to explore weaknesses in our research practices and methodologies (e.g., Bernard et al. 2004), including the quality and discrimination of our research publication outlets. The best research syntheses share the qualities of the best primary investigations. They are objective, transparent, precise, and repeatable.

By adopting the Argument Catalogue approach Abrami et al. (this issue) intended to go beyond a typical meta-analysis, while maintaining as much as possible the strengths of systematic reviews. In most respects the Argument Catalogue follows the same procedure as a meta-analytical review, including: 1) an exhaustive systematic literature search oriented towards some broad and yet applied (valuable for scientific, trade or policy practice) research question; 2) carefully defined set of comprehensive exclusion criteria consistently applied across all studies identified through searches, and 3) a coding system, developed through both theoretical analysis and review of a representative sample of studies, that is capable of accounting for a variety of reported outcomes and, no less importantly, explaining these outcomes in terms of study design and implementation (i.e., its methodological, pedagogical, and other related study features).

What distinguishes the Argument Catalogue from a meta-analysis is that the catalogue allows for inclusion of evidence from sources other than quantitative empirical research. An Argument Catalogue also attempts to account for the opinions and concerns of other parties interested and involved in e-learning—educators (teachers, administrators, and policy makers), researchers, students, even the general public presuming that public attitudes are adequately reflected in mass media).

Another difference, and a potential weakness of the Argument Catalogue, is that because of the diversity and breadth of sources, the outcomes it considers are defined more loosely, and they include not only measured results of some e-learning practices, but also their perceived impacts. Nevertheless, we saw the Argument Catalogue as an opportunity to enrich our understanding of the issues concerning e-learning by reconciling different points of view. Our study also included a quantitative summary of evidence, which reported effect sizes for 17 Canadian primary empirical research studies.

The importance of combining evidence and viewpoints.In our e-learning review, we took a broad definition of e-learning predictor variables, which included information and communication technologies used to support interactions for learning, and a broad definition of impacts (outcome variables) representing seven categories of outcomes including: achievement; motivation/satisfaction; interactivity/communication; meeting social demands; retention/attrition; learning flexibility; and cost. We also combined these outcomes (except for cost) into a single impact factor. Additionally, we distinguished between perceived and measured outcomes. The essence of this question, about the (causal) relationship between two variables, is a form of "what works" question. Moreover, we recognize the importance of distinguishing different forms of e-learning and different

outcome measures, and this was the central focus for study features coding and analysis.

However, reviews are hardly necessary, and additional research is not called for, when the findings concerning a core question are identical from study to study, or more generally, across all sources. It is the inconsistency of findings that require additional research and why reviews are undertaken. Reviews are also conducted to answer questions about the general impacts of treatments, if any, and the circumstances under which the effects vary. So in our review we also explored factors or study features which might explain the variability in findings. The essence of this phase of a review, exploring the variability in findings, is a form of exploring why things "work" along with the contextual and other factors which come into play to moderate the basic relationship between two variables. In doing so, we disagree with Anderson's (this issue) claim that it may not be valid to combine studies from disparate contexts. This concern about mixing "apples and oranges" by combining a range of studies in a systematic review has been around for some time. It is in their similarities and differences that we can identify underlying processes and note which mechanisms are generalizable and under what circumstances. Otherwise, we are left with no hope of ever being able to cumulate our understanding or apply it widely.

While we are sensitive to the sharp dichotomy of views which characterize the methodological paradigm wars between quantitative and qualitative researchers, we believe there should be more agreement between the camps about both the purpose (i.e., the why) of a single investigation and the value of synthesizing investigations, even as the debate about the form (i.e., the how) of inquiry continues. And we also hope that the debaters come to see the value of mixed methods of inquiry, like the Argument Catalogue, that combine both quantitative and qualitative methods. Finally, we believe that the best quantitative studies are well suited for hypothesis testing or confirmatory purposes about "what works." In contrast, the best qualitative studies are well suited for hypotheses generation or for exploring why things work.

This is especially the case at the level of the individual study. However, the special advantages of a systematic review are that the accumulation of evidence across studies also allows for the exploration of process explanations of why things work. It is why we believe that a good systematic review does more than cumulate what is known but adds to knowledge by exposing consistencies and inconsistencies in findings across contexts and exploring why variability exists.

Strengths and weaknesses of the Argument Catalogue. In a meta-analysis, only quantitative studies are cumulated and analysed so that effect sizes are extracted and subjected to statistical analyses for heterogeneity and model fitting. Methods like the Argument Catalogue are attempts to include a broader range of evidence into the systematic review process. Doing so means exploring the consensus of evidence across multiple sources by tallying "votes" for each outcome—positive, neutral, or negative. By this method, we observe the direction of an effect but not the magnitude of effect. Vote counts of evidence have been criticized for favouring large sample studies because, all

other things equal, a significant effect is easier to find in a large sample study than a small sample study. By extension, a consistent effect observed in a vote count is not the same as a large effect. For example, concluding that the evidence favours a positive effect of e-learning is not the same as saying the effect is small, medium, or large. "How much" questions cannot be answered by vote counts. But the cruder vote counting metric meant that we were able to include multiple sources of evidence rather than only data from quantitative primary studies. In addition, for the small number of experimental (7) and quasi-experimental (10) primary studies we located and included, we computed 29 effect sizes for the e-learning composite measure. The mean effect size was small (+0.117), but the variability among the effects was heterogeneous.

Haughey (this issue) suggests:

> For a good literature review we expect a wide gathering of possibly relevant articles, a sifting by type, a review of methods, of conclusions and a subsequent grouping by some series of constructs so as to illuminate the reader about the topic, the issues which have already been identified by previous researchers, their limitations, and possible issues which still need exploration. Does the adding of a rating scale make the process more rigorous? I don't believe so. My hope that we might have found an alternative way to explore the findings of both post-positivist and interpretivist or critical studies has not been confirmed by what I read. Perhaps so many of their findings seem reasonable because they followed the steps of a good literature review. (p. 116).

Our review contained a wide gathering of relevant articles, five different types of articles including both qualitative and quantitative empirical studies, a review of methods (e.g., highlighting the lack of field experiments) and of findings, and a subsequent grouping by constructs (e.g., the importance of instructional design). And it did not rely just on a rating scale, as we also coded articles qualitatively and summarized their major messages.

Our team reviewed all open-ended (emergent) coding intended to identify, when possible, the authors' principal positions, arguments presented, or conclusions reached, in every document. The most salient, interesting, informative, or powerful message or the most representative or frequently appearing message in each document was extracted. These were summarized and organized by the source of evidence.

Nevertheless, in Abrami, Bernard and Wade (2006), we discuss the strengths and weaknesses of the Argument Catalogue further. In our initial efforts at creating an Argument Catalogue, we attempted to strike a balance between the extraordinary breadth of the e-learning review and the depth we wanted to achieve in conducting a detailed analysis of the documents. While we covered five distinct sources of evidence, we did not do so comprehensively for all of them, choosing to focus especially on Canadian primary research studies, literature reviews and policy documents.

The Argument Catalogue codebook for our e-learning review was developed through an emergent approach by taking a representative sample of documents of various types to ensure that the major issues covered by the documents were reflected. The documents from each of the five theme areas were subsequently coded using the common codebook. The codebook can be found at the CSLP website at <http://doe.concordia.ca/cslp/CanKnow/eLearning.php>. One advantage of a common

codebook is that it allowed us to analyze the coded information from different resources within a single database, while also allowing for different types of literature to be analyzed as subsets. An example of this is the capacity to analyze study features (such as publication date or technology addressed) across all types of publications, while being able to select studies containing quantitative data to extract effect sizes for a meta-analytic review.

Throughout the process of coding information from the different resources, certain issues became evident. The more comprehensive the codebook became in order to address the different sources, the larger it grew. On one hand, this led to a higher level of overlap and interconnection between some codes. On the other hand, some codes were totally irrelevant for some resources. For example, all practitioners' articles had missing codes for research design, effect sizes, and other features that are only pertinent to primary research or reviews.

In future, we will explore the strengths and weaknesses of a more detailed and in-depth analysis of all the sources of evidence. This can be achieved through developing a more comprehensive codebook, with additional study features, including sets of features that apply specifically to each source of evidence.

Questions about cost effectiveness and value are worth considering given the time and resources needed to conduct a thorough examination of the multiple perspectives of an issue. Will a small random sample of media, practitioner and policy documents be a legitimate proxy for an inclusive search? Is the quality in the detail or is a cursory examination of the global issues sufficient? Answers to questions of this sort may only come with experience gained from seeing the extent to which, for example, systematic reviews are improved and the impact of both Argument Catalogues and reviews increases.

Our expertise and experience with the quantitative techniques, often referred to as meta-analysis, coupled with their popularity, led us to use this approach to systematic reviews of research in our initial Argument Catalogue. As the field of systematic reviewing matures, other methods are emerging including those that synthesise evidence using qualitative techniques, and Dixon-Woods et al. (2005) critically summarise a range of methods including narrative summary, thematic analysis, grounded theory, meta-ethnography, meta-study, realist synthesis, Miles and Huberman's data analysis techniques, content analysis, case survey, qualitative comparative analysis and Bayesian meta-analysis. The notions we explore here on synthesising multiple sources, not only quantitative and/or qualitative research evidence, can also be extended to these other review techniques.

In our inaugural Argument Catalogue we did not weight the primary evidence by sample size but instead treated each study equally. Similarly, we did not account for the size and scope of the documents in aggregating evidence from these sources. A medium-sized newspaper article was given the same weight as a lengthy literature review. In future, we might consider giving more weight to documents as a function of their scope.

Our analyses of evidence from the various sources were based on frequency or vote counts of impacts. We combined all sources of evidence, regardless of methodological quality, in order to have an idea about the consistency of the effect of e-learning as reflected by the different sources. Vote counts provide such information about the consistency of effects and not their size, but size matters for policy decisions, as do considerations of cost. Future research should explore both.

In most systematic reviews, especially quantitative syntheses of primary evidence, considerable effort is expended in judging the quality of the evidence using a plethora of methodological criteria often focusing on, but not necessarily limited to, a study's internal validity or the certainty with which causal inferences are likely. We did not apply these quality judgments as exclusion criteria (given the limited number of studies which would meet rigorous standards), nor did we apply any quality criteria whatsoever in judging the other sources of evidence. We did, however, compare the primary evidence, the literature reviews and the conclusions we extracted from other sources to identify similarities and differences. We understand that a review based on an Argument Catalogue gives voice to popular perceptions and the *Zeitgeist* of current views, whether they are formed carefully or carelessly. We believe that the advantages of a comprehensive review such as ours are worth the risks and the costs of broad inclusivity. But the final judgment lies with readers.

In addition, we do not know for which areas it is appropriate to undertake an Argument Catalogue and for which areas it is not. Certain aspects of health and social policy, including educational practices, seem ideal candidates for integrating the arguments from diverse constituencies. Other aspects of policy may be less ideal, for example, regarding specific pharmacological or surgical interventions. In general, an Argument Catalogue is best employed in those situations where knowing what works needs to be tempered by knowing how others perceive what works from their differing viewpoints.

With regard to the applicability of an Argument Catalogue approach, we do not yet know whether, and to what extent, it will serve its intended purpose: to inform and engage policy makers, practitioners and the general public so that more evidence is given greater credibility and exposure, and is eventually taken up and used.

We like what Schwier (this issue) suggested about the use of narrative as a complementary approach to quantitative research as a means to provide insight and develop research questions. It reminded us that academic reviews need to be translated into clear, concise language and jargon-free narratives or "stories" that policy-makers and practitioners can understand. We developed *Knowledge Links* for this purpose and prepared English and French versions for our e-Learning review. See <http://doe.concordia.ca/cslp/RA-Themes_CanKnow.php>.

Finally, our conclusions depend on the quality of our coding, which is dependent on the quality of the reports, and the reports are static snapshots of long-term complex and dynamic processes. Quantitative reviews are limited in the nature and the amount of information that can be coded, and we have no consensually acceptable techniques for

producing qualitative reviews (which are also derived from simplified reports). Given the complexity of the processes being studied, and the variability of the methods being used in primary studies, and especially the other sources of evidence we included, there are limitations to what can be coded and what can be concluded. As Bernard et al. (2004) lamented, reviews are limited by the quality, comprehensiveness and detail of the evidence they synthesize and face the challenges of what we described as the "methodological morass."

What is e-learning and what are its impacts?

Haughey (this issue) makes an articulate case that the CCL definition of e-learning we used in our review is very expansive and raises questions about the variety of possibilities it entails. Both Haughey (this issue) and Kanuka (this issue) are concerned that the breadth of the definition, and consequently the breadth of our review, mask important details because the granularity is of the review is too large.

In our review, we focused on the impacts of e-learning in general across a range of outcomes including achievement, motivation/satisfaction, interactivity/communication, meeting social demands, attrition/retention, learning flexibility, a composite impact measure, and cost. We also examined the impacts of e-learning for five CCL theme areas except the workplace: adult education, early childhood education, elementary/secondary education, postsecondary education, and health and learning. And we did not search specifically for informal learning environments as suggested by Schwier (this issue). In addition, for the data as a whole and also separately for the literature reviews and primary Canadian research combined, we looked to see whether there were facets or aspects of e-learning related to impacts including: contexts of technology use, technology tools, pedagogical uses of technology, and location.

We agree with Haughey (this issue) that the "larger societal context" should be considered whenever the impacts of e-learning interventions are addressed. The broad set of study features in our Argument Catalogue was intended to do exactly that, to the extent the data reported in the existing literature allowed. Our sincere hope is that researchers designing and conducting new studies in the area will pay more attention to investigating, and reporting on, the "societal context" in which e-learning events occur..

We produced a rough sketch of the evidence or an overall picture of recent Canadian evidence along with four other sources. This rough sketch should not be interpreted to mean there is uniformity or consistency in the findings. The generally positive impact of e-learning does not mean that all applications of e-learning have positive results, all the time, and in every context. In fact, our analyses attempted to explore the inconsistencies we found.

Table 8 of Abrami et al. (this issue) presents the summary analyses of perceived e-learning impacts on seven outcomes from five source of evidence, with the impact measured recoded so that +1 signifies a positive outcome, 0 a neutral outcome, and -1 a negative outcome. For only one of 35 results was the data uniformly positive—general

public opinion of learning flexibility afforded by e-learning. For all the other results the findings indicted some degree of variability, and we attempted to explore this variability further by analysing study features.

Similarly, we did not treat e-learning only as an encompassing term but also looked at specific features. For example, we coded for: 1) Context of technology use (i.e., distance education, in class, blended, unspecified); 2) Type of tools used (i.e., internet/intranet/on-line/web, virtual reality/learning objects/simulations, technology integration—computers and software for particular purposes, unspecified); and 3) Intervention type, including: Instructional (e.g., drill, practice, tutorials, remediation); Communicative (e.g., e-mail, ICQ, computer conferencing, LCD projector); Organizational (e.g., data base, spreadsheets, record keeping, lesson plans); Analytical/Programming (e.g., statistics, charting, graphing, drafting, robotics); Recreational (e.g., games); Expansive (e.g., simulations, experiments, exploratory environments, brainstorming); Creative (e.g., desktop publishing, digital video, digital camera, scanners, graphics); Expressive (e.g., word processing, on-line journal); Evaluative (e.g., assignments, portfolio, testing); Informative (e.g., Internet, CD-ROM); and Unspecified/Missing.

We concluded in our review that a more in-depth analysis is called for. We agree with the reviewers on this point, and we appreciate the suggestions they offered for exploring the literature further. However, there are two important caveats. First, if the collection of evidence included in our review is not extended, further analyses will not change the general results but may only explain the variability in different ways. Second, detailed examinations may show that much evidence is missing, and fine-grained analyses may conclude that more evidence needs to be collected.

Kanuka (this issue), for example, suggested looking more closely at achievement outcomes and contrasting: a) lower-level learning impacts (i.e., surface learning) and higher-level learning impacts (i.e., deep learning); b) cognitive, affective and psychomotor impacts (although we did examine achievement impacts and motivation/satisfaction); and c) disciplinary differences (e.g., language arts, social sciences, math and engineering, etc.). Similarly, she suggests a more detailed conceptualization and fine-grained analysis of the outcomes included.

Kanuka (this issue) is especially concerned about our treatment of the interactivity/communication outcome suggesting, among other things, that we should have separated interacting with technology or content from interacting with peers or the instructor. She makes a strong case by resorting to research that largely falls outside the scope of our investigation. This highlights the previous point we made about needing to find a balance between breadth and depth in any review.

Schwier (this issue) citing Downes also has a collection of learning outcomes worthy of further exploration, including: how to predict consequences; how to read; how to distinguish truth from fiction; how to empathize; how to be creative; how to communicate clearly; how to learn; how to stay healthy; how to value yourself; and how to live

meaningfully. This list reminds us of CCL's Composite Learning Index and the Conference Board of Canada's Learning Skills Profile. All these lists are conceptually rich and broad in scope, and it would be wonderful to explore e-learning impacts in these categories. If only there was evidence available to do so.

Similarly, Schwier (this issue) elaborates on our finding concerning the importance of instructional design in e-learning. He offers suggestions for new directions for the role of instructional designers that may bear fruit in future applications of technology for learning.

Otherwise, Schwier (this issue) offer ten questions that both primary research as well as reviews of research should address. While many of these questions fall outside the purview and mandate of our review, and may not be answerable until more research is undertaken, they serve an important purpose in pointing the way towards future directions for inquiry and synthesis.

Concluding remarks

As an overall concern, we are worried that in-depth reviews will become too narrow in scope to address larger questions about impact and import, losing sight of the forest for the trees. At the same time, we share the concern that too general a picture may mask underlying and important variability in impacts. The answer is that no single review, like no single investigation, answers all the questions about a topic. And like a good primary investigation, a good review should stimulate further investigations and further reviews. We hope that our review and the commentaries have done just that.

Finally, we want to thank again the Editor of the *Canadian Journal of Learning and Technology* for affording this visibility to our work. And we especially appreciate the reviewers who in critiquing our review took the task as seriously as they did. We can think of no better way of ending our rejoinder than by echoing Anderson's (this issue) point that we need to ensure that Canadians "are able to take advantage of this most important educational development since the printed text" (p.107).

References

Abrami, P.C., & Bernard, R.M. (2006). Research on distance education: In defense of field experiments. *Distance Education, 27*(1), 5–26.

Abrami, P.C., Bernard, R.B., & Wade, C.A. (2006). Affecting policy and practice: Issues involved in developing an argument catalogue. *Evidence & Policy: A Journal of Research, Debate and Practice, 2*(4), 417–437.

Abrami, P.C., Bernard, R.M., Wade, A., Schmid, R.F., Borokhovski, E., Tamin, R., Surkes, M., Lowerison, G., Zhang, D., Nicolaidou, I., Newman, S., Wozney, L., & Peretiatkowicz, A. (2007). A review of E-learning in Canada: A rough sketch of the evidence, gaps and promising directions. *Canadian Journal of Learning and Technology. 33*(1)*, 1–68.

Abrami, P. C., Cohen, P., & d'Apollonia, S. (1988). Implementation problems in meta-

analysis. *Review of Educational Research, 58*(2), 151–179.

Anderson, T. (2006). A response and commentary to: A review of e-learning in Canada.*Canadian Journal of Learning and Technology, 32*(3)*,* 101–109.

Angeli, C., Valanides, N., & Bonk, C. J. (2003). Communication in a web-based conferencing system: The quality of computer-mediated interactions. *British Journal of Educational Technology, 34*(1), 31–43.

Aviv, R., Zippy, E., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks, 7*(3). Retrieved December 30, 2005, from http://www.sloan-c.org/publications/jaln/v7n3/v7n3_aviv.asp

Bernard, R.M., Abrami, P.C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L. Wallet, P.A., Fiset, M., & Huang, B. (2004). How does distance education compare to classroom instruction? A Meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379–439.

Bernard, R. M., & Naidu, S. (1990). Integrating research into practice: The use and abuse of meta-analysis. *Canadian Journal of Educational Communication, 19*(3), 171–195.

Bonk, C., & Cunningham, D. (1998). Searching for constructivist, learner-centered and sociocultural components for collaborative educational learning tools. In C. Bonk & K. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse*. New York: Erlbaum, (pp. 25–50).

Bullen, M. (1999). Participation and critical thinking in online university distance education.*Journal of Distance Education,13*(2). Retrieved April 20, 2006, from http://cade.athabascau.ca/vol13.2/bullen.html

Cook, T. D. (2004). *Randomized experiments in education: Why are they so rare?* (Institute for Policy Research Working Paper WP-02-19). Evanston, IL: Northwestern University.

Davis, M., & Rouzie, A. (2002). Cooperation vs. deliberation: Computer mediated conferencing and the problem of argument in international distance education. *International Review of Research in Open and Distance Learning 3*(1). Retrieved September 1, 2006, from http://www.irrodl.org/content/v3.1/davis.html

De Laat, M. (2001). *Network and content analysis in an online community discourse. CSCL-ware in practice*. New York: Kluwer Publications.

Dixon-Woods, M., Agarwai, S., Jones, D., Young, B. & Sutton, A. (2005). Synthesising qualitative and quantitative methods: a review of possible methods. *Journal of Health Services Research and Policy*, *10*(1), 45–53.

Feenberg, A. (n.d.). *The "TEXTWEAVER" – Active reading hypertext for computer conferencing.* Retrieved September 22, 2005, from http://www-rohan.sdsu.edu/faculty/feenberg/textweaver/hyper.html

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education, 15*(1), 7–23.

Gunawardena, C., Carabajal, K., & Lowe, C. A. (2001). *Critical analysis of models and methods used to evaluate online learning networks.* (ERIC Document Reproduction Service No. ED456159).

Gunawardena, C., Lowe, C., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 395–429.

Haughey, M. (2006). Commentary of e-learning review. *Canadian Journal of Learning and Technology, 32*(3)*,* 111–118.

Jeong, A. (2004). *The effects of communication style and message function in triggering responses and critical discussion in computer-supported collaborative argumentation.* Paper in conference proceedings of the Annual meeting of the Association of Educational Communications & Technology, Chicago, IL.

Kanuka, H. (2005). An exploration into facilitating higher levels of learning in a text-based Internet learning environment using diverse instructional strategies. *Journal of Computer Mediated Communication, 10*(3). [online]. Retrieved September 1, 2006, fromhttp://jcmc.indiana.edu/vol10/issue3/kanuka.html

Kanuka, H. (2006). A review of e-learning in Canada: A rough sketch of the evidence, gaps and promising directions: A commentary. *Canadian Journal of Learning and Technology, 32*(3)*,* 85–100.

Kanuka,, H., & Anderson, T. (1998). On-line social interchange, discord, and knowledge construction. *Journal of Distance Education, 13*(1), 57–74.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Lopez-Islas, (2001). A cross-cultural study of group processes and development in online conferences. *Distance Education, 22*(1), 85–121.

McKlin, T., Harmon, S. W., Evans, W., & Jones, M. G. (2002). Cognitive presence in web based learning: A content analysis of students' online discussions. *IT Forum, 60*.

McLaughlin, C., & Luca, J. (2000). *Cognitive engagement and higher order thinking through*

computer conferencing: We know why but do we know how? Retrieved September 1, 2006, from http://www.lsn.curtin.edu.au/tlf/tlf2000/mcloughlin.html

Meyer, K. A. (2003). Face-to-face versus threaded discussions: The role of time and higher-order thinking. *Journal of Asynchronous Learning Networks, 7*(3), 55–65.

Nussbaum, M., Hartley, K., Sinatra, G. M., Reynolds, R. E., & Bendixen, L. D. (2002).*Enhancing the quality of on-line discussions*. New Orleans, LA: Paper presented at the annual meeting of the American Educational Research Association.

Pawan, F., Paulus, T. M., Yalcin, S., & Chang, C-F. (2003). Online learning: Patterns of engagement and interaction among in-service teachers. *Language Learning & Technology, 7*(3), 119–140. [online]. Retrieved on October 14, 2006 fromhttp://llt.msu.edu/vol7num3/pawan/

Pena-Shaff, J. (2005). Asynchronous online discussions as a tool for learning: Students' attitudes, expectations, and perceptions. *Journal of Interactive Learning Research, 16*(4), 409–430.

Pena-Shaff, J., Martin, W., & Gay, G (2001). An epistemological framework for analyzing student interactions in computer-mediated communication environments. *Journal of Interactive Learning Research, 12*, 41–68.

Pena-Shaff, J., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in Computer Bulletin Board (BBS) discussions. *Computers and Education, 42,*243–265*.*

Rourke, L. (2005). *Learning through online discussion*. Unpublished Ph.D. Dissertation. University of Alberta,, Edmonton, Alberta, Canada.

Rovai, A., & Barnum, K. (2003). Online course effectiveness: An analysis of student interactions and perceptions of learning. *Journal of Distance Education, 18*(1), 57–73.

Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Koschmann, (Ed.), *CSCL: Theory and practice of an emerging paradigm*(pp. 249–268). Mahwah, NJ: Erbaum.

Schwier, R. A. (2006). Commentary on "A review of e-learning in Canada". *Canadian Journal of Learning and Technology, 32*(3)*,* 71–83.

Thomas, M. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning, 18,* 351–366.

Vaughan, N., & Garrison, D. R. (2005). Creating cognitive presence in a blended faculty development community. *Internet and Higher Education, 8*(1), 1–12.

Veerman, A., Andriessen, J., & Kanselaar, G. (2000). Learning through synchronous

electronic discussion. *Computers & Education, 34*(3–4), 269–290.

Wilson, D., Varnhagen, S., Krupa, E., Kasprzak, S., Hunting, V., & Taylor, A. (2003). Instructors' adaptation to online graduate education in health promotion: A qualitative study. *Journal of Distance Education, 18*(2), 1–15.

Yakimovicz, A., & Murphy, K. L. (1995). Constructivism and collaboration on the Internet: Case study of a graduate class experience. *Computers in Education, 24*(3), 203–209.