Journal of Information Systems Education

Volume 30
Issue 1
Winter 2019

The Role of Flow in Learning Distributed Computing and MapReduce Concepts using Hands-On Analogy

Colin Conrad, Michael Bliemel, and Hossam Ali-Hassan

Recommended Citation: Conrad, C., Bliemel, M., & Ali-Hassan, H. (2019). The Role of Flow in Learning Distributed Computing and MapReduce Concepts using Hands-On Analogy. *Journal of Information Systems Education,* 30(1), 57-66.

Article Link: http://jise.org/Volume30/n1/JISEv30n1p57.html

Initial Submission:	27 January 2018
Accepted:	20 September 2018
Abstract Posted Online:	4 December 2018
Published:	13 March 2019

ISSN: 2574-3872 (Online) 1055-3096 (Print)

The Role of Flow in Learning Distributed Computing and MapReduce Concepts using Hands-On Analogy

Colin Conrad
Rowe School of Business
Dalhousie University
Halifax, NS B3H 4R2, Canada
colin.conrad@dal.ca

Michael Bliemel
Faculty of Business and Information Technology
University of Ontario Institute of Technology
Oshawa, ON L1H 7K4, Canada
mb@uoit.ca

Hossam Ali-Hassan
International Studies and Business Administration
York University
Toronto, ON M4N 3M6, Canada
hossama@glendon.yorku.ca

ABSTRACT

The expansion of technical concepts into everyday business practices suggests a need for effectively teaching difficult subjects to non-technical users. This paper describes hands-on analogy, an innovative method for teaching technically difficult concepts using interactive, experiential learning activities and a gamified exercise. We demonstrate our technique by investigating Hadoop Hands On, an exercise designed to teach MapReduce. Students experienced how MapReduce functions work conceptually by envisioning students as compute and tracking nodes in a Hadoop system and playing cards as data processed to complete two tasks of varying complexity. A study of 56 students was conducted to validate the exercise and demonstrated the impact of triggered flow on perceived understanding. The main contributions of this work are 1) an alternative learning approach that communicates a technically difficult concept through analogy and 2) the demonstration of the role of flow in facilitating learning using this approach. We recommend using this approach to teach technically difficult concepts to non-technical students who can more easily comprehend the benefits of distributed computing methods interactively in a way that complements the traditional lecture approach.

Keywords: Active learning, Analogy learning, Game-based learning, Big data

1. INTRODUCTION

The pedagogy of Management Information Systems (MIS) involves teaching complex technical concepts to students who may have varying degrees of interest or familiarity with a given subject. As the advancement of new business technologies accelerates, so does the demand for teaching complex concepts quickly. Many new teaching techniques in MIS have sought to evoke intrinsic motivation among the students by triggering flow or cognitive absorption (Agarwal and Karahanna, 2000; Léger et al., 2010), especially in the development of collaborative, serious games that use simulation. These approaches have exhibited promising results in the case of teaching technical skills in the management classroom (Cronan et al., 2012) and have demonstrated broad effectiveness for knowledge acquisition in other settings (Boyle et al., 2016).

However, MIS instructors have to teach technologies or concepts that are not only technical in nature, but also technically unfeasible given the target audience. For instance, students may learn how to manage IT infrastructure projects, but acquiring the technical skills necessary to run an IT infrastructure simulation is prohibitive for most business students. These problems are particularly pronounced when working with Big Data infrastructure, as it is technically complex and increasingly essential knowledge for business school graduates (Phillips-Wren et al., 2015). An alternative approach is to teach using interactive hands-on analogy. Unlike direct computer-based simulation, hands-on analogy has the benefits of eliciting learning without having to learn complex task-specific skills and can be delivered through a non-complex medium. Analogy has been successfully implemented in the teaching of science and mathematics (Treagust and Duit, 2015) and can help facilitate inferential learning with respect to complex subjects (Niebert, Marsch, and Treagust, 2012). Further, by using a low-technology medium to deliver an analogy, we can further control for extraneous challenges characteristic of the technological multimedia through which serious games are usually delivered. The process of teaching through interactive analogy thus draws attention by triggering enjoyment or flow without distractions. We theorize that hands-on analogies that trigger flow can be used to teach complex technical concepts effectively and efficiently, in a way that is appropriate for students pursuing a business or managerial education.

To test our approach, we created a technique called *Hadoop Hands On*, an exercise for teaching students about the MapReduce algorithm using playing cards. MapReduce is an algorithm for performing distributed computing tasks using clusters of computers. It has played a central role in the expansion of Big Data especially through the proliferation of the Apache Hadoop open source platform (Dean and Ghemawat, 2008). Hadoop and MapReduce are used by hundreds of companies, including Twitter, Amazon, and Yahoo! for processing the large datasets that are becoming increasingly essential to the operations of large businesses (Stonebraker, 2014; Connolly, 2015). Understanding the technical details of MapReduce often requires knowledge of databases and algorithms characteristic of those taught to senior computer science students. The *Hadoop Hands On* technique exemplifies our approach by using hands-on analogy to trigger flow.

In this paper, we describe our theory of hands on analogy and its effectiveness in common IS education contexts. We then describe the *Hadoop Hands On* technique and a study of 56 business students who went through the exercise. The students were asked to provide their perceived knowledge of the subjects of MapReduce and Hadoop before taking part in the exercise. After taking part in the exercise, the students were again asked to report their perceived knowledge of MapReduce, and the approach was found to be effective. Finally, we describe future research that can be conducted on the role of hands-on analogy in Management Information Systems education using psychophysiological measures.

2. HYPOTHESIS DEVELOPMENT

Serious games are a growing interest in the subject of management education. Agarwal and Karahana (2000) outlined a Cognitive Absorption construct, which has been established as a flow measure in the Management Information Systems literature. Games have been shown to be effective for engaging students and teaching complex concepts in IT Management (Bliemel and Ali-Hassan, 2014). Recent findings by Lu, Hallinger and Showanasi (2014) confirmed that simulations offer the potential for dramatically improving the quality of university-based management education. Additionally, deeply engaging games and student experiences are not restricted to electronic media. Forming small, intentional groups for informal cooperative learning is an effective method to trigger engagement in an engineering classroom (Smith and Sheppard, 2005). Recent developments in "reverse classrooms," where students do not participate in lectures but instead participate solely in active learning, have been found to be at least as effective for producing strong grade point averages as traditional classroom environments (Baepler, Walker, and Driessen, 2014).

The primary motivations for teaching using hands-on analogy are to a) provide alternative learning tools that communicate technically difficult concepts to non-technical audiences and b) to provide intrinsic motivation for learning about an otherwise difficult or boring topic. Analogy and metaphor have been studied in the field of science education, particularly in its role to change children's conceptual framework governing scientific phenomena. Building on their earlier work on conceptual change (Duit and Treagust, 2003), Treagust and Duit (2015) describe the role that metaphor can play in bridging abstract concepts and schemas that make up students' understanding of perceived phenomena. This is described well in their analysis of the use of analogy to teach energy in Physics (Lancor, 2014). For example, by teaching that energy is like money, which can be accounted for (as opposed to energy being an abstract property), students can begin to bridge the abstract with the familiar. Students may start to demonstrate their understanding of energy by describing a series of coherent analogies, but as their abstract understanding further develops, students often find more accurate ways of describing perceived phenomena using the abstract terms.

However, when teaching a topic such as an algorithm to business students, analogy may not be sufficient, as students are often not intrinsically motivated to learn about subjects outside of their primary experience. An alternative approach is to design a vehicle for the analogy that is intrinsically motivating by triggering flow. Since its original conception by Csikszentmihalyi in the 1970s, flow has been an important concept in learning, particularly in the literature related to serious games. Serious games describe games that are used to achieve a teaching objective. Flow describes the enjoyable state of being intensely absorbed in an activity, where users are motivated to engage in an activity for its own sake, rather than for extrinsic reasons (Csikszentmihalyi, 1990). If serious games are able to trigger flow, students are intrinsically motivated to engage in the activity, and it is thus easier to keep the learner's attention. Much of the appeal of serious games is their ability to provide an alternative tool for educators to teach concepts that would otherwise be difficult, such as in the case of a technically complex subject. The literature on the subject of serious games has further pointed toward the potential for these tools to educate in Science, Technology, Engineering, and Mathematics (STEM) disciplines, given the role that perceptual skills play in these fields (Connolly, 2012; Boyle et al., 2016).

In order to determine whether participants experience flow, we first need an appropriate flow measure. Finneran and Zhang (2005) conducted a survey of flow measures and the challenges faced with their implementation, with attention to the IT training context. Notably, the authors indicated a need for flow measures to be adapted to the dynamic nature of the experience, which is something that survey measures had thus far been limited at achieving. Recent research on the topic of flow and cognitive absorption have yielded robust psychophysiological measures that can overcome the real-time challenge (Léger, 2014), however these can be deemed inappropriate for the classroom environment.

Pearce and Howard (2004) sought to overcome the limitations of questionnaire instruments by constructing a flow process measure that consists of a simple ratio of perceived challenge and skill. The two-question measure contains two five-point Likert scale questions inspired by Csikszentmihalyi's original conception of flow. Pearce, Ainley, and Howard (2005) validated this measure with information systems students who participated in an e-learning exercise designed to teach students topics in physics. Pearce, Ainley, and Howard (2005) also employed an ex-post flow measure that included other constructs related to Csikszentmihalyi's conception, such as control, enjoyment, and engagement. By doing this, they could investigate the relationship between a simple two-question flow-state measure with a more robust flow conceptualization.

However, these findings also faced two significant challenges. The first was the reliability of the measure: how do we know that students made similar judgements when evaluating skill? Different students might evaluate the same objective skill level differently. More challenging however is the discrepancy between the flow process measure and a more robust flow-state measure, collected ex-post. After performing factor analysis, Pearce, Ainley, and Howard (2005) ultimately conclude the investigation by acknowledging these limitations, but by recognizing the value of the overall state-level measure of flow and its effectiveness in learning contexts. Their ex-post flow-state measure used an 11-item instrument, but dropped 2 of these items following factor analysis. The resulting nine-item questionnaire was able to account for 64.6% of the variance and detected significant differences between students who had more mastery of the subject material versus others. We thus conclude that asking participants questions related to their experienced flow-state (e.g., "I found the exercise enjoyable") is an effective way to measure flow in teaching contexts. This leads us to articulate our experimental hypotheses:

H1: Participants in the hands-on analogy exercise will perceive attaining knowledge from the exercise.

H2: There is a positive relationship between the experienced flow-state and the perceived attainment of knowledge from the exercise.

3. RESEARCH DESIGN

To demonstrate the effectiveness of our technique, we created the *Hadoop Hands On* exercise and conducted a study of students who participated in the exercise. The study aimed to discover whether students learned Hadoop MapReduce concepts, whether they experienced flow, and whether there is a relationship between the flow experienced and learning. This section describes the exercise in detail, the chosen survey instrument, and the collection procedure.

3.1 Description of Hadoop Hands On Exercise

The *Hadoop Hands On* exercise is designed to reinforce concepts covered in undergraduate and MBA Business Analytics and Management Information Systems classes. The format of these classes at our university uses a flipped classroom approach where students are responsible for reading and learning the class concepts prior to arriving to class where these are then reinforced through mini lectures, discussion, and hands on experiential learning in our teaching labs. Our Business Analytics courses run over 13 weeks and cover topics such as decision support, data acquisition, data preparation, data modelling, data cubes, business reporting, and text analytics, before exploring Big Data concepts. For our classes on Big Data, students were expected to read a textbook chapter on the subject which included three pages on MapReduce and Hadoop. During the class, we reviewed related Big Data concepts and then taught Hadoop and MapReduce concepts using the *Hadoop Hands On* exercise. These concepts were relevant for future lessons related to data mining and emerging trends in business analytics, as well as a hands on module using predictive analytics software. Students were later quizzed on material from the topics and also had the option to utilize these concepts as part of a final portfolio project. In addition to the interactive component, *Hadoop Hands On* includes a 10-minute PowerPoint presentation which described the origins of the MapReduce algorithm and described the exercise. A summary of the key elements of the presentation are provided in the following subsection.

3.2 What is MapReduce?

MapReduce was originally developed by Google in the early 2000s to process large amounts of raw data, such as crawled internet documents, using clusters of low cost computers (Dean and Ghemawat, 2008). The computation itself takes a set of user-defined inputs and produces a set of outputs according to the specified task and contains two primary functions: map and reduce. The map function is used to identify and sort the data according to the user specifications. The reduce function performs a computation, such as counting or aggregating. Rather than computing on a single machine, MapReduce uses clusters of inexpensive, commodity computers to perform tasks. Figure 1 describes the high-level operation.

When a user calls the MapReduce function, the user triggers a multi-step process invoking the nodes of the cluster. The program begins by splitting the input files into manageable sizes which are then assigned to various "worker" machines by a special "master" node. The master node then assigns map and reduce tasks to the workers. Workers assigned with map tasks proceed to identify data. As the map workers make progress, the master node notifies reduce workers of the location and nature of the processed data. The reduce workers iterate over the sorted data and eventually pass the results of the reduce function to the master node, completing the MapReduce call.

The Hadoop framework is a popular rendition of the MapReduce algorithm. Initially conceived in the mid-2000s (Abouzeid et al., 2009), Hadoop is open source software (The Apache Software Foundation, 2015) maintained by the Apache Foundation. Hadoop is optimized for commodity hardware and maintains advanced routines accounting for failures. In addition to Hadoop's original MapReduce program, Hadoop contains a number of other features that advance its functionality. Together with these features, Hadoop has extended the MapReduce function. Today, Hadoop is the standard for processing data from heterogeneous sources and has become a supplement to traditional data warehousing technologies.

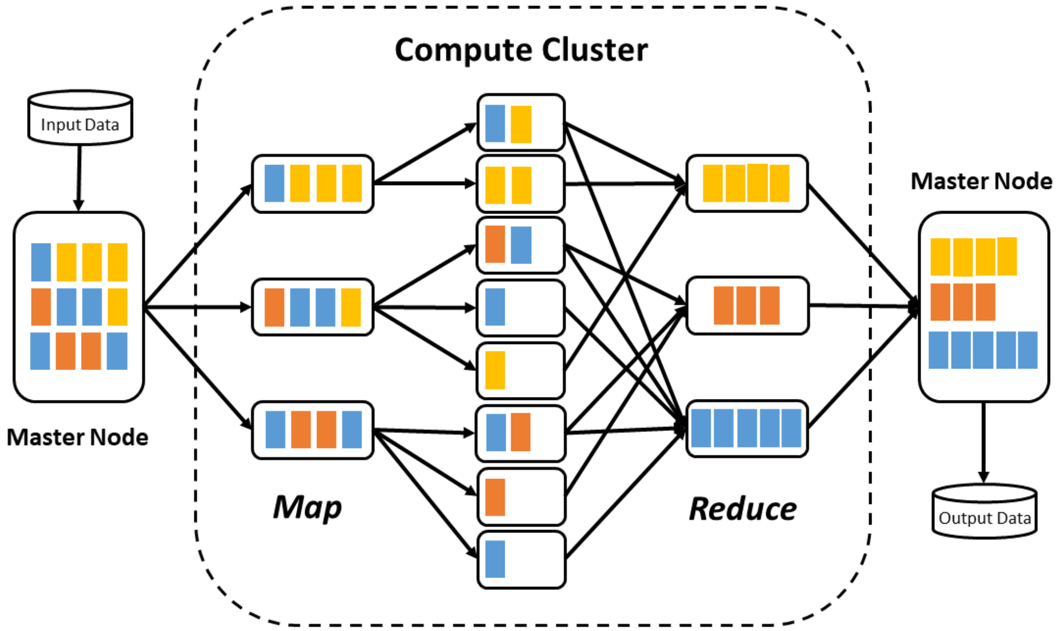

Figure 1. The Process of MapReduce

3.3 Exercise Instructions

We introduced these MapReduce concepts in a mini lecture to students that was reinforced with the *Hadoop Hands-On* activity. The analogy is to view the class as the compute cluster in a Hadoop Distributed File System. The materials required to complete this are several decks of standard playing cards as well as name tags printed for students taking on management roles. The class was split into two teams that competed against each other, where each team represented a Hadoop system of multiple compute clusters. One student in each team was given a name tag of Job Tracker who then led each of the two teams, or Hadoop systems. Name tags of Task Tracker were handed out to the leader of each of the sub groups or clusters. Each cluster consisted of a subgroup of three to five students. The remaining students without nametags represented the worker nodes that do the data processing. Teams were encouraged to compete with each other to complete the task as quickly as possible. Competition facilitates focused interest, cognitive absorption, and challenge, which are constructs characteristic of flow.

Each of the two Hadoop systems / teams of students was given six shuffled decks of playing cards. The data that was processed in this file system is represented by these playing cards. Here we continue the analogy by explaining to students that these cards represent the text and numbers from Amazon product reviews by many different people.

The scenario we worked with was that there were four snow shovels to choose from labeled by suit – Hearts, Spades, Clubs and Diamonds. Each card is one review that identifies the product (represented by the suit) and the score, represented by a number 2-10. Non-number cards (Ace, King Queen, Jack, Joker, Instruction cards, etc.) represent irrelevant text data, which can be ignored in our first activity. This type of data is relatable to students, who are then instructed that the goal of the exercise is to sift through large amounts of data for different products.

In the first part of the *Hadoop Hands On* exercise the question was asked "Which product has the best reviews?" The solution for this is found by counting the points of each suit. To randomize the problem, an arbitrary 10-20 set of cards was removed from the 6 shuffled decks. The data was then given to the team leader / Job Tracker who was instructed to fairly distribute the data to sub-group leaders / Task Trackers, who in turn distribute the work to their Worker Nodes. The Map Process was then initiated, where each worker node maps the data by product type and review score or sum of all the points of a suit. This was accomplished by sorting the cards into five piles (numbered cards for Hearts, Spades, Diamonds, and Clubs, and Other Cards such as A, K, Q, J, Jokers, Instruction cards). In the left side of Figure 2, we demonstrate how the cards are distributed from the Job Tracker to the Task Trackers and Worker Nodes. This is similar to how the Master Node distributes work in the Map process depicted in Figure 1.

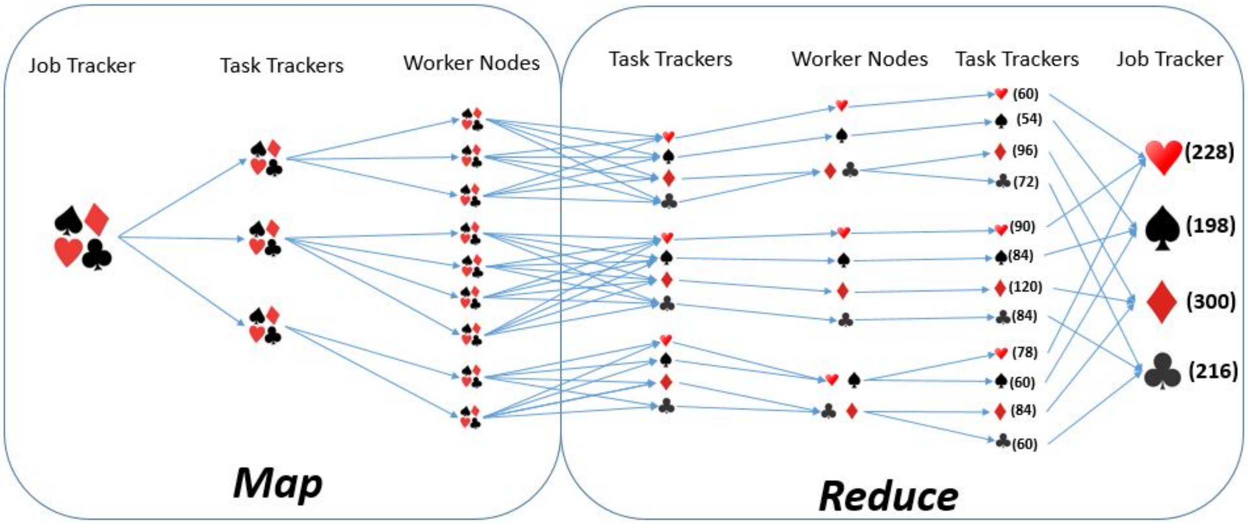

Figure 2. The Process of Hadoop Hands On

The Reduce process began once every Worker Node in a group has sorted their cards. In this process, the Task Trackers move data across nodes so that all the cards are combined for each suit. Depending on the number of nodes in a group, this could be one or two suits per student. Worker Nodes then added up the number of points for each suit and reported these to the Task Tracker. The Job Tracker asks each Task Tracker to report their totals for each product and then combined these to determine the highest rated product (suit). This is then verified by examining the cards removed at the beginning of the exercise to see which suit has the most points. This process is described by the right side of Figure 2 where the cards are sorted and ultimately transferred to the Job Tracker, similarly to the Reduce function depicted in Figure 1. Our studies suggest that this exercise takes between 5 and 11 minutes to complete depending on the group structure and the efficiency of its members.

After shuffling all the cards together again, the class then works on a more challenging scenario which reinforces concepts of distributed computing through the application of the analogy concepts to a new problem. We further increased involvement by challenging the class to solve the second problem as quickly as possible while sharing times from other classes and sections. In the second scenario we randomly picked one card from the six decks without revealing it and tasked the class with discovering which card was chosen. The Task Trackers then met with the Job Trackers to determine what strategy they should employ to solve this problem in the most time efficient way possible whilst following the rules. Developing strategies encourages deliberate and strategic thinking as well as a sense of control over the exercise. We then gave the data (cards) to the Job Trackers and began timing the process for our Hadoop Cluster to find the missing card. Our studies suggest that this part of the exercise takes between 3 and 11 minutes to complete.

During the exercise we witnessed situations where one Task Tracker was not properly fulfilling his role and seemed confused and falling behind. The Job Tracker removed his name tag and attached it to a Worker Node, effectively transferring the role to a more efficient and effective node. In another case one Worker Node was slower than the other nodes in its cluster so the Task Tracker transferred some of the cards to another node. In a third case, the Task Tracker of a smaller cluster took on the duties of a Worker Node for efficiency purposes. Following the exercise, participants take part in a debriefing session designed to reinforce the concepts explored in the analogy. In addition, concepts such as fault tolerance, workload balancing, and horizontal scaling are explored while referring to examples from the exercise analogy. The total time for the exercise ranges from 25 to 35 minutes when the introductory instruction and debriefing are included.

3.4 Instrument Design

The Pearce, Ainley, and Howard (2005) ex-post flow-state measure consisted of 11 five-point, Likert survey questions used to measure control, interest, and enjoyment. These included the following items: control, absorption, enjoyableness, thinking of other thoughts, interest, frustration, boredom, distraction, curiosity, knowing what to do, and concentration. As previously mentioned, Pearce and Howard (2004) also included a simple two-question measure of challenge and skill in their investigation which was included in our study as control variables, in addition to the 11 questions. A measure of perceived understanding was developed by our team, which consists of differences in perceived understanding of the learning objectives pre and post exercise. The measures consisted of understanding the following items: cluster computing, the role of Job Tracker, the map process, the reduce process, the role of Task Tracker, and why we use MapReduce. A list of the questions asked are provided in Appendices A and B.

3.5 Procedure and Data Collection

Business students who were otherwise attending the *Hadoop Hands On* exercise through their Business Analytics course were invited to take part in the survey. Fifty six students consented to participate in the study. Participants were asked to complete a pre-session questionnaire which contained information about their perceived understanding of topics such as the map and reduce algorithms, cluster computing, the use of

Variable	Mean (before)	SD before	Mean (after)	SD after	t-test significance (1-tailed, paired)
I understand the Map process of Hadoop MapReduce	1.47	1.09	5.49	1.26	0.000***
I understand the Reduce process of Hadoop MapReduce	1.45	1.02	5.56	1.14	0.000***
I understand the concept of cluster or parallel computing	2.45	1.38	5.71	0.98	0.000***
I understand why we use Hadoop MapReduce	1.55	1.10	5.64	1.44	0.000***
I understand the role of the Task Tracker in Hadoop MapReduce	1.31	0.88	6.05	1.18	0.000***
I understand the role of the Job Tracker in Hadoop MapReduce	1.25	0.78	6.04	1.09	0.000***
Average	1.58	1.04	5.75	1.18	
Note: *p<0.05; **p<0.01; ***p<0.001					

Table 1. Comparison of Reported Understanding Variables Pre and Post Exercise

Variable	Mean (before)	SD before	Mean (after)	SD after	t-test significance (1-tailed, paired)
I understand the Map process of Hadoop MapReduce	3.14	1.22	5.29	1.60	0.026*
I understand the Reduce process of Hadoop MapReduce	3.00	1.41	5.71	7.56	0.003**
I understand the concept of cluster or parallel computing	3.14	1.34	5.43	0.98	0.006**
I understand why we use Hadoop MapReduce	3.29	1.80	5.57	1.72	0.040*
I understand the role of the Task Tracker in Hadoop MapReduce	1.86	1.22	5.57	1.72	0.002**
I understand the role of the Job Tracker in Hadoop MapReduce	1.86	1.22	5.43	1.72	0.002**
Average	2.72	1.37	5.5	2.55	
Note: *p<0.05; **p<0.01; ***p<0.001					

Table 2. Comparison of Reported Understanding among Participants with Prior Knowledge of MapReduce

map reduce when solving business problems, and the role of task or node trackers. During the activity, task completion times were recorded. Following the activity, students were asked to complete a questionnaire which included the flow instrument and the same questions about perceived understanding. The differences in perceived knowledge were calculated and saved as a delta measure. Data from one of the participants was removed due to incomplete responses.

4. ANALYSIS AND RESULTS

Of the 55 respondents, 16 were females and 39 males; 20 were enrolled in the Bachelor of Commerce, 2 in the Bachelor of Management, 15 in the Master of Business Administration (MBA), 8 in the Master of Library and Information Studies (MLIS), and 10 in the Master of Electronic Commerce (MEC) program. When asked about having "any previous knowledge of Hadoop MapReduce," 48 students answered "no" and 7 said "yes." During the experiment, 5 students had the role of Job Tracker, 14 were Task Trackers, and the remaining 36 were Worker Nodes. We shall now revisit the hypotheses in light of the respondents' results.

H1: Participants in the hands-on analogy exercise will perceive attaining knowledge from the exercise.

In order to assess the overall learning of the students, we compared their responses to the six questions about their general understanding of MapReduce before the exercise with their responses after the exercise. The average understanding before the exercise was 1.58 out of 7 with a standard deviation of 1.04. This indicates that students did not properly understand the concepts from the textbook readings which described the algorithm over three pages using text and diagrams alone. After the exercise, the average level of understanding went up to 5.75 with a standard deviation of 1.18. A paired t-test (one-tailed) on the six outcome variables before and after the game were all significant with $p<0.001$ (Table 1), confirming that students who participated in the exercise ended up understanding the concept of MapReduce.

In addition to the overall learning effect, we can observe differences among the seven participants who reported prior knowledge of Hadoop MapReduce. Although the sample is small, we can observe that participants with prior knowledge reported higher mean responses before the exercise, but still reported significant increases in knowledge afterward. Though the mean knowledge reported was higher than the average, it was below 4 which suggests that participants who reported prior knowledge on the subject perceived their knowledge was limited. This suggests that the exercise may be appropriate for participants with some degree of understanding, in addition to participants who report no prior understanding. Table 2 summarizes the results from the seven participants who reported "any previous knowledge of Hadoop MapReduce." Given that students reported an increased understanding of the MapReduce concept, Hypothesis 1 is clearly supported. Participants clearly perceived attaining knowledge from the exercise.

H2: There is a positive relationship between the experienced flow-state and the perceived attainment of knowledge from the exercise.

Variable	Descriptive Statistics: Mean (SD)	Beta	Std. Error	p-value
Absorption	5.84 (1.10)	0.04	1.13	0.787
Boredom	1.89 (1.26)	0.22	0.94	0.158
Concen-tration	6.36 (0.89)	0.18	1.15	0.178
Curiosity	5.56 (1.23)	0.02	0.88	0.874
Distraction	4.09 (1.67)	0.08	0.46	0.400
Enjoyable-ness	5.93 (1.03)	-0.06	1.19	0.683
Control	5.22 (1.30)	-0.46	0.75	0.001**
Frustration	2.76 (1.80)	0.02	0.49	0.878
Interest	5.46 (1.11)	0.39	0.86	0.002**
Thinking other thoughts	2.76 (1.77)	-0.03	0.66	0.821
Know what to do	5.40 (1.16)	0.44	0.74	0.000***
Challenge	3.64 (1.77)	-0.08	0.46	0.420
Skill	5.91 (0.97)	0.44	0.99	0.001**
Note: *p<0.05; **p<0.01; ***p<0.001				

Table 3. Results from Multivariate Linear Regression Test between Flow and Reported Understanding

To study the role of flow in understanding the concept of MapReduce during the exercise, we conducted a multivariate linear regression test to examine the relationship between the flow construct and the outcome variable, Reported Understanding. For this analysis, the outcome variable was defined as a composite change in the perceived understanding measures (listed in Table 1) and was calculated by subtracting the reported values from before the exercise from the respective reported values after the exercise. Descriptive statistics for the independent variables (mean and SD) are listed in Table 3 along with the regression coefficients and significance values of each of the 11 flow items and the two control items.

Regression analysis using the flow measure resulted in an R-square of 0.693, while ANOVA analysis demonstrated an F-value that was significant at $p<0.001$. Analysis on individual items revealed that four items are significantly associated with Reported Understanding: control, interest, knowing what to do, and skill. To test for multicollinearity, the VIFs between all the independent variables included in the model were calculated and the highest value found was 3.26, a sign of low risk of multicollinearity. The fit between the flow measure and Reported Understanding supports our second hypothesis.

5. DISCUSSION

In examining the Reported Understanding measures, we found that students ended up with significantly increased perceived understanding of the different concepts around distributed computing and MapReduce. In fact, their average perceived understanding increased from 1.58 before the exercise to 5.75 afterwards. We also found support for our hypothesis that there is a positive relationship between the experienced flow-state and the perceived attainment of knowledge concerning distributed computing, albeit partially. The flow measure used was associated with the perceived attainment of knowledge, but of the 13 items that constituted the measure, only 4 were significantly associated with the outcome variable. Though the measure was able to account for 69.3% of the perceived learning, we expected that more of the variables would significantly contribute to the variance.

Pearce, Ainley, and Howard (2005) performed factor analysis to identify common factors from the questions and found that items could be explained by two factors, which they would label "enjoyment" and "control." One significant variable observed in our study (Interest) can be identified with the former, while two significant variables observed in our study (Control, Know what to do) can be identified with the latter. One potential issue is that the control item was significant but was negatively associated with the attainment of learning outcomes. It is possible that students who were too focused on the mechanics of the exercise reported high degrees of control and also reported less understanding. In addition, one additional variable (Skill) was significantly associated with reported learning. It is possible that by including Skill, we managed to capture a dimension of flow that was not captured well by the original flow-state measure described by Pearce, Ainley, and Howard (2005).

In adopting this exercise to explain distributed computing and MapReduce, it is important for instructors to provide clear instructions to students on performing the assigned roles and manipulating the cards in the mapping and reduce phases of the exercise. It is also important for students to feel that the exercise does not require skills that they may lack. When explaining a technical concept like MapReduce to business students, it is best to use plain English and avoid technical terminology. Finally, in the future, when instructors are using analogy to explain other technical concepts with new exercises, they should make sure the exercises are interesting for students, that they are tailored to their skillset, and that the provided instructions are clear.

As discussed, one of the primary challenges of working with an ex-post flow measure is that it is unclear whether this measure accurately accounts for the changes in experienced flow throughout the exercise. By having a real-time measure of flow, we could offer insight into the variances of experience among the participants. A second significant limitation to our results is that though the model is sound, the study measures a relationship between flow and perceived learning as opposed to an objective learning measure. Implementing an objective measure that accurately reflects the learning objectives is challenging but a necessary component of a comprehensive result. A third limitation of this study is the relatively small sample size. It is possible that the effects of additional flow items would have been observed with a larger sample.

6. CONCLUSION

The idea of using hands-on analogy to teach technical concepts to non-technical students shows promise. The *Hadoop Hands On* exercise demonstrated an engaging, low-technology tool for teaching business students about distributed computing and the MapReduce algorithm. These topics are technically complex and difficult to teach to students who lack technical training. Using analogy and an interactive game, students experience flow, and when they experience flow it helps them form a better understanding of the subject. This exercise does not require extensive training to implement, is technology agnostic, and can complement a traditional IS education module on big data and distributed computing. Given the short duration of the lesson, this could be a useful tool for teaching non-technical students about the MapReduce algorithm.

In addition, we have demonstrated an instrument for measuring flow and its impact on perceived learning in such an environment. Though this measure is administered ex-post, it is able to account for the relationship between perceived learning and the participant's subjective evaluation of their experienced flow-state during the exercise. This is a useful measure for conducting analysis of experienced flow in a classroom environment. Given that it can be administered ex-post, it has the benefit of not disrupting the classroom activity.

However, these findings are limited by the subjective nature of the flow measure and would benefit from further investigation using objective, real-time measures of learning outcomes, and in doing so, answer the concerns raised by Finneran and Zhang (2005). Psychophysiological measures might be able to adequately measure some of the elements of flow and offer a potential topic for further investigation. States such as stress or involvement may be measured using the galvanic skin response or heart rate, and there are non-intrusive devices that might be appropriate for the classroom environment. In moving this work forward, an investigation of the potential for these psychophysiological measures could yield insight into the real-time flow experienced by participants and could deliver insights into the future improvement of the exercise.

7. ACKNOWLEDGEMENTS

Colin Conrad would like to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and Killam scholarship during the process of this research. The authors would also like to acknowledge the support by a grant from the Social Sciences and Humanities Research Council of Canada (SSHRC).

8. REFERENCES

Abouzeid, A., Bajda-Pawlikowski, K., Abadi, D., Silberschatz, A., & Rasin, A. (2009). HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads. *Proceedings of the VLDB Endowment,* 2(1), 922-933.

Agarwal, R. & Karahanna, E. (2000). Time Flies when You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly,* 24(4), 665-694.

Apache Software Foundation (2015). *Apache Hadoop.* Retrieved May 15, 2015, from https://wiki.apache.org/hadoop/PoweredBy.

Baepler, P., Walker, J., & Driessen, M. (2014). It's not about Seat Time: Blending, Flipping, and Efficiency in Active Learning Classrooms. *Computers & Education,* 78, 227-236.

Bliemel, M. & Ali-Hassan, H. (2014). Game-Based Experiential Learning in Online Management Information Systems Classes using Intel's IT Manager 3. *Journal of Information Systems Education,* 25(2), 117-124.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C. & Pereira, J. (2016). An Update to the Systematic Literature Review of Empirical Evidence of the Impacts and Outcomes of Computer Games and Serious Games. *Computers & Education,* 94, 178-192.

Finneran, C. M. & Zhang, P. (2005). Flow in Computer-Mediated Environments: Promises and Challenges. *Communications of the Association for Information Systems*, 15(1), 82-101.

Connolly, S. (2015). Going from Hadoop Adoption to Hadoop Everywhere. Retrieved May 15, 2015, from http://hortonworks.com/blog/going-from-hadoop-adoption-to-hadoop-everywhere/.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education,* 59, 661-686.

Cronan, T., Léger, P.-M., Robert, J., Babin, G., & Charland, P. (2012). Comparing Objective Measures and Perceptions of Cognitive Learning in an ERP Simulation Game. *Simulation and Gaming,* 43, 461-480.

Csikszentmihalyi, M. (1990). *Flow and the Foundations of Positive Psychology.* New York, NY: Springer.

Dean, J. & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM,* 51(1), 107-113.

Duit, R. & Treagust, D. F. (2003). Conceptual Change: A Powerful Framework for Improving Science Teaching and Learning. *International Journal of Science Education*, 25(6), 671-688.

Lancor, R. A. (2014). Using Student-Generated Analogies to Investigate Conceptions of Energy: A Multidisciplinary Study. *International Journal of Science Education*, 36(1), 1-23.

Léger, P.-M. (2014). Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context. *Computers in Human Behavior*, 34, 273-283.

Léger, P.-M., Davis, F. D., Perret, J., & Dunaway, M. M. (2010). Psychophysiological Measures of Cognitive Absorption. *Proceedings of SIGHCI 2010.*

Lu, J., Hallinger, P., & Showanasai, P. (2014). Simulation-Based Learning in Management Education. *The Journal of Management Development,* 33(3), 218-244.

Niebert, K., Marsch, S., & Treagust, D. F. (2012). Understanding Needs Embodiment: A Theory-Guided Reanalysis of the Role of Metaphors and Analogies in Understanding Science. *Science Education,* 96(5), 849-877.

Pearce, J. & Howard, S. (2004). Designing for Flow in a Complex Activity. *Proceedings of the 6th Asia-Pacific Conference on Computer Human Interaction,* 349-358.

Pearce, J., Ainley, M., & Howard, S (2005). The Ebb and Flow of Online Learning. *Computers in Human Behavior*, 21(5), 745-771.
Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business Analytics in the Context of Big Data: A Roadmap for Research. *Communications of the Association for Information Systems,* 37*,* 448-472.
Smith, K. A. & Sheppard, S. D. (2005). Pedagogies of Engagement: Classroom-Based Practices. *Journal of Engineering Education,* 94(1), 87-101.
Stonebraker, M. (2014). Hadoop at a Crossroads? Retrieved May 15, 2015, from http://cacm.acm.org/blogs/blog-cacm/177467-hadoop-at-a-crossroads/fulltext.
Treagust, D. F. & Duit, R. (2015). On the Significance of Conceptual Metaphors in Teaching and Learning Science: Commentary on Lancor; Niebert and Gropengiesser; and Fuchs. *International Journal of Science Education*, 37(5-6) 958-965.

AUTHOR BIOGRAPHIES

Colin Conrad is a Lecturer in Information Systems at Dalhousie University in Halifax, Canada. Colin is also currently a Ph.D. candidate in Interdisciplinary Studies where he studies through Dalhousie's Rowe School of Business, Department of Psychology and Neuroscience, and Faculty of Computer Science. His research interests are in human-computer interaction, cognitive science, and education technology.

Michael Bliemel is a Professor in Management Systems and Dean of the Faculty of Business and Information Technology at the University of Ontario Institute of Technology in Oshawa, Canada. He has diverse interdisciplinary research experience around the impacts of new technologies, data literacy, business analytics, e-health, gamification, information systems, e-commerce, human computer interaction, and the innovation and adoption of emerging technologies.

Hossam Ali-Hassan is an Assistant Professor of Business Administration at Glendon College, York University, in Toronto. He holds a Ph.D. in MIS from Schulich School of Business. His current research interests include business analytics, data literacy, experiential learning, social media, social capital, and job performance. Prior to his academic career, he worked for many years as a technologist and consultant.

Appendix A: Flow Measures Adapted for the Exercise

a) During the exercise, I was totally absorbed in what I was doing [Absorption]
b) The exercise bored me [Boredom]
c) I was concentrated fully on the exercise [Concentration]
d) The exercise excited my curiosity [Curiosity]
e) During the exercise, I was aware of distractions [Distraction]
f) I found the exercise enjoyable [Enjoyableness]
g) During the exercise I felt in control of what I was doing [Control]
h) I was frustrated during the exercise [Frustration]
i) The exercise was intrinsically interesting [Interest]
j) I thought about other things during the exercise [Thinking of other thoughts]
k) I knew the right thing to do during the exercise [Knowing what to do]
l) How challenging did you find the exercise? [Challenge]
m) Were your skills appropriate for understanding the exercise? [Skill]

Appendix B: Perceived Understanding Measures Administered Before and After the Exercise

a) I understand the concept of cluster or parallel computing
b) I understand the role of Job Tracker in Hadoop MapReduce
c) I understand the Map process of Hadoop MapReduce
d) I understand the Reduce process of Hadoop MapReduce
e) I understand the role of the Task Tracker in Hadoop MapReduce
f) I understand why we use Hadoop MapReduce

Information Systems & Computing
Academic Professionals

EDSIG
Serving Information Systems Educators

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN 2574-3872