Journal of
Information
Systems
Education

Volume 30
Issue 2
Spring 2019

Teaching Tip
"The Data Shuffle": Using Playing Cards to Illustrate Data Management Concepts to a Broad Audience

David Agogo and Jesse Anderson

Recommended Citation: Agogo, D. & Anderson, J. (2019). Teaching Tip: "The Data Shuffle": Using Playing Cards to Illustrate Data Management Concepts to a Broad Audience. *Journal of Information Systems Education,* 30(2), 84-96.

Article Link: http://jise.org/Volume30/n2/JISEv30n2p84.html

Initial Submission:	24 September 2018
Accepted:	14 January 2019
Abstract Posted Online:	13 March 2019
Published:	5 June 2019

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

Teaching Tip
"The Data Shuffle": Using Playing Cards to Illustrate Data Management Concepts to a Broad Audience

David Agogo
Department of Information Systems & Business Analytics
Florida International University
Miami, FL 33199, USA
dagogo@fiu.edu

Jesse Anderson
Big Data Institute
Reno, NV 89501, USA
jesse@jesse-anderson.com

ABSTRACT

Educators must constantly figure out engaging ways to teach data management and modeling concepts, especially to non-technical audiences. This paper introduces and describes an experiential learning activity using playing cards to teach a range of business and technical concepts. The paper is enriched by personal anecdotes and experiences from conducting this activity in both academic and professional settings. A repeated measures survey (pre-test, post-test, and follow-up one week later) is used to evaluate the effectiveness of the exercise. Participants reported enjoying the exercise, demonstrated improved understanding, felt confident about their new knowledge, and recalled important concepts a week later.

Keywords: Experiential learning & education, MapReduce, Playing cards, Metaphors in IT education, Data management

1. INTRODUCTION

Due to the abstract nature of data management concepts and related subjects, teaching these concepts to an audience with only basic technical knowledge requires careful planning and technique. Early generations of digital information storage were skeuomorphic (i.e., modeled against physical forms of storing information), making them easier to understand. For instance, hierarchical file storage was consistent with physical file drawers (Harrison, 2015). However, modern databases are highly abstract, employing advanced relational concepts and more complicated data structures designed to meet the use cases and realities with which today's businesses must contend. Such use cases include storing and analyzing big data and managing streaming data from social media and connected devices. Instructors in the field of information systems (IS) must deal with this challenge and figure out innovative ways to pass along course content while keeping students engaged and interested. In this paper, one such approach to teaching important concepts from this knowledge area is introduced.

At the heart of this approach is an Experiential Learning Activity (ELA) which leverages the notion of reverse-skeuomorphism (when "digital-things" or ideas are re-applied to "physical things" (Scott, 2012)). The physical objects used in this ELA are a regular deck of playing cards. The goal of this ELA is to improve learners' understanding and retention of important data management concepts and make the teaching of these concepts more effective. By using playing cards, this exercise reuses a tool with which participants are likely to be familiar. This paper is important and relevant because it equips IS instructors with a useful, effective, and fun exercise that they can deploy immediately in the classroom. Learners also stand to benefit, as the use of this exercise is likely to encourage students' further engagement with more advanced concepts. At the very least, this ELA is shown to make the knowledge and understanding of data concepts more accessible to a wider audience, including those who are less technically inclined.

In the following section, a brief review of the use of ELAs to teach important data and programming concepts is provided. Thereafter, the exercise is introduced in detail. Discussion of the exercise will cover planning for the activity, conducting the activity in a classroom setting, and suggested guidelines for debriefing. Finally, the paper presents empirical evidence that this approach is effective and concludes with a discussion of the empirical findings and participant experiences.

2. LITERATURE REVIEW

2.1 Using Experiential Learning Activities to Teach Concepts

Experiential learning activities are activities that facilitate the creation of knowledge through the transformative effect of experience (Kolb, 1984). Instead of the traditional practice where students passively sit through lectures or read text, "the learner is directly in touch with the realities being studied… rather than merely thinking about the encounter or only considering the possibility of doing something with [the knowledge obtained]" (Keeton and Tate, 1978, p.2 in Kolb, 1984). ELAs have been used in a wide variety of fields, especially when practical applications are important (e.g., in business or design) and when the concepts are particularly difficult to master. In the field of IS, ELAs have been used to teach design thinking (Gaskin and Berente, 2011) and IT consulting (Heim et al., 2005), among other concepts. These activities range from custom-made simulation environments to more flexible and ad-hoc thought exercises using the most basic of tools. However, the embeddedness of technology in today's classroom and the presence of web 2.0 technologies tend to spur and support a range of ELAs (Huang and Behara, 2007).

A key challenge with the design and delivery of ELAs is the amount of work that goes into planning and crafting a meaningful ELA experience for learners. Building online simulations and tools requires a significant amount of work and may end up being unnecessarily rigid; therefore, approaches that utilize simple and easy to understand tools are particularly useful. For instance, ELAs have used marshmallows and sticks of spaghetti (Cook and Olson, 2006), colored beads (as in the famous red bead experiment by Edward Deming, and subsequent adaptations) (Gartner and Naughton, 1988; Turner, 1998), colored envelopes and tokens (Hamey, 2003), Lego bricks (Ammar and Wright, 1999), and playing cards (Enders, Stuetzle, and Laurenceau, 2006; Hakulinen, 2011), among others. This paper focuses on the use of playing cards to teach important data management concepts.

The idea of using playing cards in an ELA is not new. For instance, a regular deck of playing cards has been used to teach qualitative analysis, random assignment, data structures and algorithms, and software engineering processes. Waite (2011) teaches how tacit theory and explicit theory shape qualitative data analysis by deconstructing the way participants sorted a shuffled set of playing cards when no instructions were provided. Another group has used playing cards to illustrate how random assignment works and how spurious patterns may emerge despite randomization (Enders, Stuetzle, and Laurenceau, 2006). This was done by asking students to randomly assign a shuffled deck of playing cards into a predetermined number of groups, tally the occurrence of 'background variables' (e.g., the number of cards of a certain color, number, or shape) between groups, then perform T-tests on the patterns observed. Yet another educator, Hakulinen (2011), describes the use of a deck of playing cards to teach data structures and algorithms as part of a computer science course. The challenge of sorting a shuffled deck of cards is used as a context to illustrate the properties of algorithms, explain stable versus in-place sorting methods, and evaluate the effectiveness of simple algorithms. A very ambitious project (Baker, Navarro, and Van Der Hoek, 2003, 2005) uses a custom set of trading cards to explain the software development process, walking participants through requirements gathering, design, module building and integration, and maintenance according to the waterfall software development process. These educators created a deck of custom trading cards, with each card having scenarios and details that participants had to respond to at each stage of the development process. Participants competed to gain the most points which were awarded based on how well they responded to each scenario that emerged during gameplay.

2.2 Teaching Data Management Concepts using ELAs

Past work that reports the use of ELAs to teach data management and database concepts tend to describe complex online simulations and animated courseware. For instance, to teach database concepts, Connolly, Stansfield, and McLellan (2006) designed a game-based, online learning simulation based on constructivist ideals, encouraging students to take control of their own learning within an authentic, realistic, and sufficiently complex problem space. The researchers report on positive results from giving students access to this simulation, including more enjoyment and better performance. However, students complained about the workload and found it difficult to manage their time during the simulation (Connolly and Begg, 2007). Another project developed animated courseware to teach and test on a range of database concepts (Murray and Guimaraes, 2008). This learning aid also included an interactive SQL query builder, allowing students to visualize the resulting information at each stage of building a query. Evaluation of this tool found positive effects such as increased depth and breadth of database knowledge, increased testing performance, and generally positive evaluations from students and faculty alike (Guimaraes, 2006). However, as attested to by the papers describing these projects, these online tools are complex and difficult to build, maintain, and manage. There is an opportunity to create and disseminate simple and effective ELAs using readily available tools such as playing cards. The following section describes an exercise designed for this purpose.

3. AN EXPERIENTIAL LEARNING ACTIVITY & TEACHING FRAMEWORK

3.1 Data Management Curriculum and Learning Goals

Learning about data management spans a range of important topics, such as an introduction to data types and data modeling; data storage; information retrieval; relational database principles; and creating, updating, querying, and modifying databases. While covering this curriculum, learners are introduced to abstract concepts that diverge from a commonsense understanding of how physical information is stored. This often leads to confusion and difficulties in comprehension. Further, there are even more advanced concepts that are required to learn to work with data effectively. A working understanding of indexes, views, constraints, normalization of tables, and query optimization are all important knowledge. Further, given the evolution of next-generation NoSQL databases, a new generation of data management systems different from the familiar tabular formats further increases the level of complexity and abstractness. Wholistic instruction on databases must also cover these NoSQL databases including column databases, key-value pairs, document databases, and graph data. This myriad of approaches to data management are further complicated by the increasing popularity of distributed storage and processing of information, made possible by technologies such as Hadoop, MapReduce, and so on. Further, the rapid generation and spread of information require an understanding of how to deal with streaming data in efficient ways. Besides introducing students to the technical aspects of these systems, it is critical for learners to understand the principles behind how each of these different systems handle the storage and processing of information. While it is certainly too ambitious for any single exercise to serve as an introduction to all these concepts, this ELA was designed keeping this body of knowledge in mind to serve as a critical introduction to these concepts. In addition, there are also important analytical and business-related concepts which are essential to a robust data management curriculum.

Modern organizations exist within a business environment that has been transformed by the presence and power of data. The success stories of organizations that have been able to respond to this new reality are regularly promoted and celebrated. However, the reality is that many other organizations are left behind. Merely seeking to use data is not enough, as most projects embarked on to improve the value obtained from organizational data tend to fail or underperform. Estimates such as a 70%-80% failure rate are commonly quoted in the trade press (Kernochan, 2011). As such, while introducing the previously listed technical concepts of databases and data management, it is critical to stimulate an appreciation for the complexity and difficulty that a business faces when seeking to utilize their data assets. In fact, this aspect of teaching about data management is even more important for students enrolled in non-technical programs, such as business information systems courses or MBAs. Outside of formal higher education, these same concepts often need to be introduced to business managers who do not require much expertise in technical details but exert influence on technology strategy decisions and therefore need to be better informed. Given this dual focus of imparting both technical and business concepts using an ELA, the following section goes into the specifics of the proposed ELA.

3.2 Learning Objectives of the ELA

The objectives of this ELA are to use playing cards to introduce students to:

1. business concepts related to the management of data
2. several technical concepts related to the management, storage, and processing of data

In the sections that follow, a description of how the ELA is being used to teach important data management concepts is presented.

3.3 Introduction and Overview

This exercise is designed to illustrate important technical and business concepts to students with little or no knowledge about data management. It is designed to be completed in groups of three to five students. Extensive prior knowledge of data management is not required to participate meaningfully. The use of a familiar and common medium is a key goal of this exercise, so prior exposure to regular playing cards is preferred. As it will be shown, this exercise seeks to leverage existing tendencies towards how playing cards are arranged to introduce new concepts. As such, the more familiar and relatable the playing cards are, the more effectively the abstract concepts being taught will be understood. If any students are unfamiliar with playing cards, make sure they are assigned to a group with other students who are willing to explain playing cards to them.

This exercise will involve the distribution of stacks of playing cards to groups of students and then a walkthrough of several competitively timed exercises which are punctuated with feedback and discussion. At the end of the entire exercise, a class discussion is recommended to address any student confusion and answer additional questions that arise. In the authors' experiences, this ELA can be a rather active and noisy affair, so it is probably well suited for an icebreaker on the first day of a class (such as a database class in which many of the concepts are to be addressed in some additional detail). The entire activity takes between 30-45 minutes to complete, including pauses for brief discussion and feedback. Depending on the level of conversation in the debriefing exercise, the total duration may be closer to 45 minutes than 30 minutes. An overview of all the stages of this ELA is shown in Figure 1.

3.4 Planning for the Activity

To prepare for this activity, the instructor should purchase and prepare multiple decks of playing cards for use in the class. There should be at least one full deck of cards for each group of students. Each deck of playing cards should be shuffled and randomized as best as possible before their use in class. When

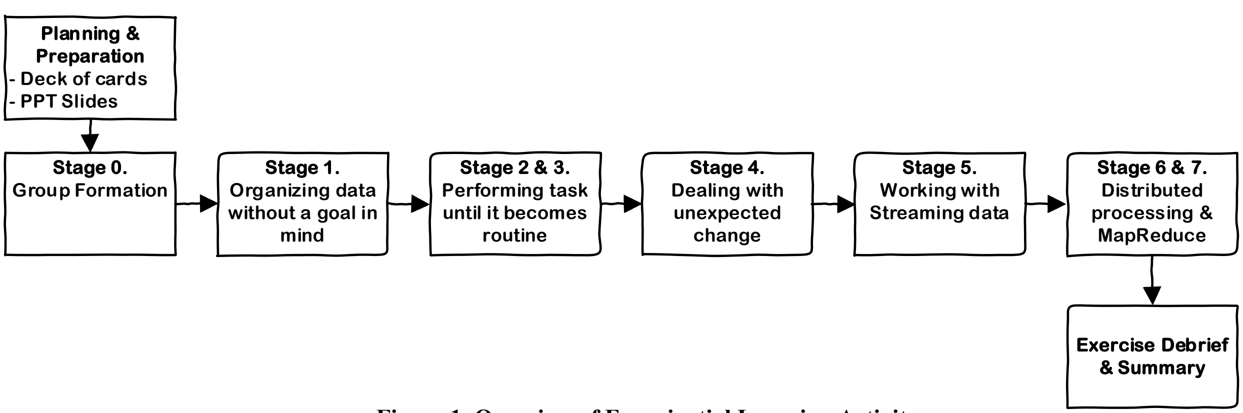

Figure 1. Overview of Experiential Learning Activity

carrying out this ELA, the authors benefitted from creating accompanying PowerPoint slides to reinforce the verbal directives and highlight the major insights from each stage of the activity. The instructor should edit and refine their set of slides (and sections of the game included) to match the specific concepts they would like to introduce. A sample subset of PowerPoint slides used by one of the authors along with a link to a complete set of sample prompts for this game are provided in Appendix A.

During the initial design of this ELA, it was anticipated that a timer readout would be needed between phases. In practice, careful timing of each group's performance on each respective phase turned out not to be so important. Also, the main concepts were still effectively imparted without needing to present actual performance times during the activity. However, the author(s) could see how others seeking to extend this exercise may benefit from either timing performance or asking groups to monitor how long they take to complete certain tasks.

3.5 Executing the Activity

The execution of the activity is reported in the following sections. The use of sections below is so that each step of the ELA can be matched with the learning objectives stated earlier. Further, the modular description allows instructors using this ELA to eliminate sections that are not relevant for their context and also to include additional exercises that they come up with. This activity is broken into three parts. In Part A (stage 0 to stage 4), groups will work with the full deck of cards. In part B, (stage 5 to stage 6), groups will also work with the full deck, but will only be allowed to retrieve a single card at a time for processing. Part C, which has to do with illustrating how MapReduce works, will require both splitting the deck of cards into a few smaller decks and retrieving a single card at a time from each smaller deck.

3.5.1 Stage 0: Group Formation and Distributing the Playing Cards. The first step is to form groups of three to five students per group. Common methods for group formation are better discussed in other work, such as Chapter 21 in Davis, 2009. As groups are formed, the instructor hands out a full deck of cards to each group with the cards face down. Groups are instructed not to pick up the cards until instructions are provided for the activity.

3.5.2 Stage 1: Now that We Have Data, What Next? The first instruction to all groups is to "look at the cards and decide how to store them." This rather vague and open-ended instruction is sure to draw some confused looks and additional questions from some groups. In response, inform the students that the point of this initial exercise is for them to figure out how they want to stack the cards unguided or unbiased by requirements. After giving a few moments (about two minutes) for this to be done, request for different groups to report on how they stacked the cards. It will be observed that groups have adopted different approaches. The key point of this stage can then be displayed, namely "businesses cannot always tell how they will use data when they collect it."

Another question that comes up during this phase is how to deal with the alphabet cards in the deck (e.g., Jacks, Queens, and Kings). Leave this question unanswered for this stage by letting groups know it is up to them to decide.

3.5.3 Stage 2: Here's Your Task. In the next stage, the groups are then informed that their task will be to retrieve one or more cards that fall between an upper and lower number limit (e.g., between 7 and 15) as quickly as possible. They are then allowed to discuss and possibly modify the way they have arranged their cards. Give the room about two minutes to rearrange the cards if necessary in preparation for the next stage. After this is done, the instructor can once again solicit responses from the room, asking if and why any groups modified their arrangements from the prior stage. The key point of this stage can then be displayed to the room, "having specific tasks in mind will bias how you store and organize organizational data."

3.5.4 Stage 3: Getting Good at Routines. Following from the instructions from the previous stage, a number falling between the two limits is called out (and displayed on the slides) with groups competing to pull up the cards as quickly as possible. This part of the exercise is likely to become rowdy. It is likely that some groups that had suitably organized their deck of cards before this stage will finish most quickly the first few rounds; however, other groups will begin to readjust their approaches between rounds. After three to four rounds of repeating this, it is expected that all groups will be much faster at retrieving cards between the set limits. The key point of this stage can then be displayed to the room, "organizations can get good at doing routine tasks with data." For discussion, students can be asked to mention some routine tasks for which organizations may utilize data.

3.5.5 Stage 4: Dealing with the Unexpected. This stage involves bringing in some unexpected realities to the game. Given that the groups spent the last few minutes getting very good at retrieving cards that fall between the lower and upper limit you set, they will not be expecting a drastic change in the rules. Ask the class to pull up cards summing to a number two to three times greater than the upper limit you had broadcasted in the previous rounds. As expected, this will draw initial shock, confusion, and maybe some audible complaints. Remain quiet through all this and, as soon as some groups begin to raise their cards upon completion, most of the other groups will begin trying to figure out the right cards as well. As soon as the commotion has quieted down, and attention returns to the front of the room (where you are most likely smiling in amusement), display the key point of this part of the exercise to the class, "when requirements or demands change for a business, past decisions on storing and managing information may become constraints." In order to salvage the mood of the room, ask groups if they would like to change the way they have organized their data or otherwise prepare better for the following rounds. Let them know the rules have changed and any number may be called. This can be repeated once or twice before proceeding to the next stage.

3.5.6 Stage 5: Dealing with Streaming Information. At the start of this stage, point out that each group had been dealing with a static situation and sorting through cards with which they were already familiar. Explain that organizations sometimes have to figure out how to analyze data as they are being generated by customers. Ask for one volunteer from each group to collect all their playing cards and shuffle the cards to demonstrate the difficulty of analyzing streaming information.

Once each group has one volunteer in possession of their cards, inform the class that the volunteer will now be moving to a new group along with the new set of shuffled cards. The instructions are that the new group member (holding the cards) can only hand out cards one at a time to the rest of the group, and the person handing out cards cannot be part of analyzing the cards any longer. The switching of groups as described guarantees that each group must analyze cards they have not seen before in real time. (An alternative approach is to ask the groups to place the cards face down on the table and to only pick one card at a time. This option is simpler to adopt and has the benefit that all group members can participate in the task. However, it is more difficult to administer in a large class or among competitive groups that are tempted to peek into the deck of cards). Once the instructions are clear, announce a target number in a similar fashion to the previous stages. Groups are to pull up cards summing to the stated number as quickly as they can. After the first round, share the key point from this stage of the exercise, "modern businesses have to deal with data that is constantly pouring in from different sources. Systems built for static data fall short with such new paradigms." At this point, allow the groups to discuss and re-evaluate how best to organize themselves to succeed at this task. After some moments of discussion, repeat the exercise again. Make sure to remind the groups that they cannot use cards that have already been handed to the rest of the group unless the entire deck of cards is reshuffled appropriately. Also, at this point, let half of the groups know they can only produce red cards and the other half can only produce black cards. At the end of repeating this exercise once or twice, ask groups to share the approaches that worked best for them.

The discussion about the approaches used may segue into discussions about the use of simple filtering or more complex algorithms to analyze large amounts of streaming data. If not, this is the appropriate point to introduce these key technical concepts to the class. In the context of this exercise, filtering is equivalent to setting a selection rule (e.g., only black or red cards) and only looking at cards that meet that rule. The behavior of summing up cards to the target value as they were received also serves to illustrate how a simple algorithm works. One common approach observed from carrying out this ELA in practice is represented below for an individual assigned to pick red cards only summing up to a 'target value' (e.g., 43). It involves keeping the target sum in mind and creating a mental record of 'card sum' which will be incremented by the value of each red card selected, provided the card was a large one (e.g., greater than 5). When the card sum neared the target value, the selection principle then changed with the individual holding out for the card that would complete the sequence to arrive at the target value. This process is pictorially represented in a flowchart form as Figure 2. The instructor can sketch out this flowchart to the class to stimulate discussion about possible

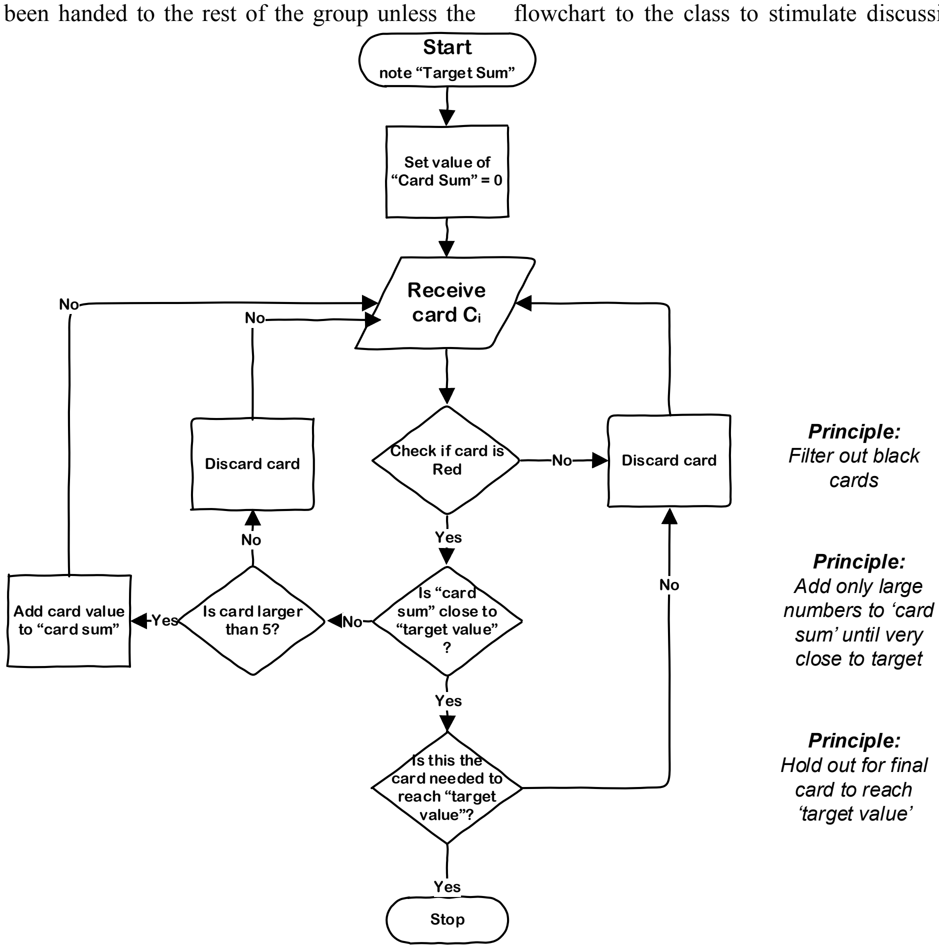

Figure 2. Flow Chart of Card Summing Approach

alternatives or ways to improve the quality of this algorithm. In a more advanced class, this point may lead to a discussion about optimization algorithms and other more advanced concepts such as neural networks which extend from this simple example.

After explaining how this algorithm works, the instructor should then generalize the idea of algorithmic thinking and how it plays a role in analyzing not just streaming data, but static data that has been stored. It is not unexpected that creating a flowchart and the rather detailed discussion of these ideas may trigger anxiety in less technologically-capable participants due to the newness and complexity encountered. The instructor can make this easier by acknowledging the reality that algorithmic thinking is quite difficult to become comfortable with and requires practice. Also, to stimulate further curiosity, the instructor may inform the class that this sort of codification of the steps to solve a problem is at the heart of artificial intelligence applications.

3.5.7 Stage 6: Distributing the Processing of Data. Next, the instructor should prompt the group to think about how to process the streaming information faster by splitting up the cards and retrieval between themselves, with each group member performing a separate operation before combining the information using a final step. Give each group some time to work on a method for carrying out this operation and report back to the entire class. The goal of this stage is to have participants thinking about the potential advantages of conducting parallel processing of information. It is likely that participants will overestimate how easily they can get themselves to work in parallel, so it is important to assign a task to allow them to try out the approach on which they decided. Remember to remind the groups to decide on how to process the information without actually sorting the cards, i.e., with cards still face down on the table as in the previous stage.

This discussion should transition into an explanation of the major challenges associated with parallelization. Some important concepts that can be discussed further include the following: (1) load balancing: the challenge of equally dividing the work between nodes in an operating cluster, (2) managing communications between nodes: dealing with the number of messages and clarification required during the actual processing of information in parallel, and (3) node failures: dealing with what happens when information assigned to one of the group members is no longer accessible, or the output of their analysis is incomplete/wrong.

3.5.8 Stage 7: Distributed MapReduce Algorithm. The final stage of this exercise is intended to illustrate how the distributed MapReduce algorithm works. This part of the exercise was first created and made popular by the excerpt of an online training (Anderson, 2013b) published on the video-sharing website YouTube (Anderson, 2013a). For this stage of the activity, the instructor may demonstrate how the algorithm works to the class first, before asking students to replicate his actions with their respective card decks. This exercise also makes a great final activity as the distributed MapReduce algorithm can be used to count the total number of cards in each suite from all the groups in the room, providing a positive and interesting final activity that shifts mindsets from competing between themselves to working together.

The first part of this stage involves an explanation of the mapping and reducing activities to the class. MapReduce is a programming model for batch processing large amounts of data in a way that can be easily parallelized to run on multiple nodes at the same time. It was made popular by Google, Inc. in a paper first published in 2004 (Dean and Ghemawat, 2004, 2008). Data is processed through several steps that are split into two major phases, a 'mapper' phase and a 'reducer' phase. After some preliminary explanation of MapReduce and its history, the instructor should then introduce an appropriate task to illustrate how MapReduce works. The task of summing up the total number of cards in each suite is one that has been used successfully (Anderson, 2013b, 2013a).

The instructor takes a stack of cards and explains that the mapper step in MapReduce involves reading in the data, line by line, and carrying out a basic operation on each card, same as the process described in the previous stages for streaming data. Cards are then read one by one and then placed in a unique stack based on their shape. That is, all hearts are placed in a single stack of hearts, all spades in a single stack of spades, and so on. After this step is completed in front of the class, the instructor explains that the reducer involves calculating the totals to report the total value for each suite. Once a basic understanding of this simple procedure is clear, the instructor now lets the class know that the distributed MapReduce algorithm involves these same steps but replicated across several nodes (from the tens to the hundreds of nodes). To deal with issues encountered in the previous stage when the groups attempted to parallelize their work with the cards, it is explained that MapReduce includes an important 'shuffling-sorting' process.

The instructor then explains that each node, under the instructions from a master node, can transfer the appropriate subset of data to other nodes so that a single node has all the relevant data it requires to work on a single suite of cards. The groups can then be asked to practice this shuffling-sorting activity within themselves using the deck of cards with which they have been working. Groups are to split all the cards among themselves – working in parallel like in the previous stage. Thereafter, each group member is to read each card in turn, sorting each new card into the appropriate stack so there is a distinct stack of hearts, spades, diamonds, and clubs only. Then, members of each group are to transfer cards between themselves until one member is left holding all the hearts, another is left holding all the spades, and so on. Finally, after this shuffle-sort is complete, each member can now sum up the cards they are holding and present the total tally of each suite as the final answer.

As a fun and possibly quite chaotic finale, all participants in this exercise may be asked to work together as a single distributed system and sum up the total numerical value of cards in circulation in the room. Groups will have to sort the cards they have and then shuffle them with other groups so that each group holds a homogenous set of cards. Where there are more groups than there are suites of cards, a most likely situation, groups can pick numbers as the sorting value instead, with group one holding all the one cards, group two holding all the two cards, and so on.

3.6 Guidelines for Debriefing

At the end of this exercise, it is important to collate all the major insights and findings and discuss them collectively. A summary slide which pulls up all the key ideas from the ELA may be useful for this purpose. Also, the instructor may paraphrase the activities and insights from some of the stages. One possible approach which may help to reinforce learning of these concepts further is the use of prompt cards that participants can fill out as the exercise progresses. An example of a prompt card with spaces that can be filled in by students is shown in Appendix A. Also, the instructor can use a summary slide to show participants each stage of the activity completed and the specific concepts or insights that the stage was designed to demonstrate. A sample table matching the stages above to concepts described is shown in Table 1. However, instructors are encouraged to treat this paper as a starting point, improving and expanding on the use of playing cards to teach other important concepts to a broad audience.

	Business Concepts
Stage 1	B1. Organizations can never anticipate all possible uses of data a priori
Stage 2	B2. Having certain tasks in mind will bias how organizations store data
Stage 3	B3. Organizations can get good at doing routine tasks with familiar data
Stage 4	B4. When data usage demands change, past decisions also amount to constraints
Stage 5	B5. Data management approaches that operate with static data fall short with new paradigms
	Technical Concepts
Stage 5	T1. Streaming data & real-time processing
Stage 6	T2. Parallel processing
Stage 7	T3. MapReduce

Table 1. Concepts Covered in each Stage of Experiential Learning Activity

3.7 Potential Pitfalls

ELAs are often plagued by different forms of dishonesty due to the potential and incentive to act dishonestly in the presence of competition and given the novelty that often comes with these activities (Schibrowsky and Peltier, 1995). This exercise is no different. In fact, anecdotal evidence reveals some potential pitfalls associated with running this exercise. For instance, it was observed that during the time groups were meant to be planning how they would store, retrieve, or analyze streaming cards efficiently, some were looking through the prior unseen stack of cards like in the previous stages. Their motivation was clearly to be the first group to announce the answer to the class. To address this, the instructor should make sure to monitor the class and point out groups that are not paying attention to the instructions. Secondly, the instructor should ensure to carry out repeat runs of the different stages where necessary so that participants are able to catch up with the instructions and stay engaged. The use of PowerPoint slides or even an interactive quiz tool (e.g., Kahoot.com) is quite helpful. More details of samples that instructors can adapt for their own use are provided in Appendix A.

4. EVIDENCE OF EFFECTIVENESS

To empirically test the effectiveness of this ELA in imparting these concepts, a repeated measures survey was conducted. Participants were given a pre-test before participating in the exercise and then two post-tests, one immediately after the exercise and the other a week after the exercise was performed (responses were collected between one and two weeks after the activity was performed). Evidence of effectiveness was evaluated by comparing within-person changes in the following measures: (1) performance on a pop quiz and (2) confidence with explaining key concepts in an exam. In addition, single-period reports of how much participants enjoyed the activity, unaided recall, and guided recall were measured.

4.1 Methods

Participants in this study were drawn from juniors and seniors enrolled in an information systems course. This sample seemed appropriate as they are one likely target group for this exercise and because they were not complete novices to the concepts covered in the exercise. Institutional review board approval was obtained for this study. The pre-test commenced by gauging participants' confidence with taking an exam in this area (e.g., If given an exam today, I would receive full credit for the question… why is data management important to a business?). Afterward, a pop quiz (comprising five true/false questions and three multiple choice questions) was administered. After this, the ELA commenced as described in the previous sections. At the end of the activity, participants once again rated how confident they were about taking an exam in this area and answered the same pop quiz questions again. A new set of questions evaluating participants' enjoyment of the exercise was also administered (e.g., I learned a lot from the card exercise, I found the card exercise quite enjoyable). Finally, after one to two weeks, participants were asked to take a follow-up survey. Participants did not have any specific data management readings or instruction between the pre-test and post-test (class content focused on an SQL coding refresher). In the follow-up survey, participants began by mentioning any insights they could remember from the ELA. This question was intended to evaluate unaided recall of the concepts taught. Afterward, a set of items measuring general recall of the exercise was administered, followed by another measurement of confidence with taking an exam in this area, and a repeat of the pop quiz from the earlier surveys. The very last question in this survey was an open-ended prompt to collect any comments or general impressions of the exercise. Details of these measures are contained in Appendix B.

4.2 Participants

Twenty five participants completed the pre-test. Of these, all but two participants completed the post-test. However, only 18 participants started the follow-up survey (with 15 responses being complete and usable). Given this difference in sample sizes across periods and the generally low sample size in this study, cases were dropped on a pair-wise basis for tests to retain as much statistical power as possible. Overall, the small sample size was not considered to be a challenge given the anticipated strength of the effect of the ELA. Actual effect sizes (Cohen's d) were also calculated for the t-tests performed.

4.3 Results and Findings

4.3.1 Performance Gains were Observed and Held Consistent during Follow-up. A paired sample t-test was performed to evaluate the changes in performance on the pop quiz across the three measurement periods. It was found that performance on the pop-quiz increased and then remained consistent during the follow-up survey. The average score of 3.91 on the pre-test increased 20 percent after the activity to 4.70 (ΔScore = +0.78, two-tailed p-value = 0.068, n = 23, Cohen's d = 0.44). Despite the marginally significant result in the two-tailed test, the increase in average score is significant if the more powerful directional one-tailed test is used (p = 0.034). Between the post-test and the follow-up, a week later, there was no significant increase or reduction in overall performance (ΔScore = +0.37, two-tailed p-value = 0.313, n = 15, Cohen's d = 0.27). The individual-level changes in performance scores for each participant are plotted in Figure 3.

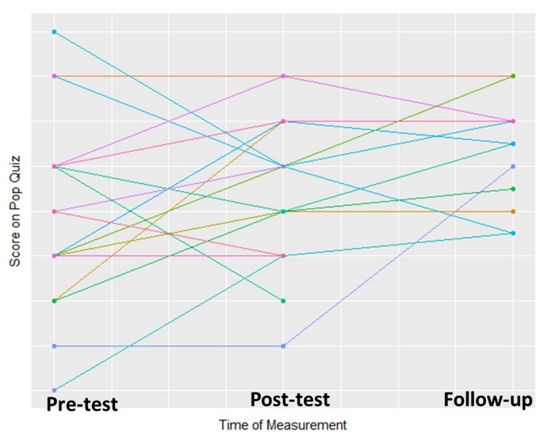

Figure 3: Line Graph of Participants' Scores on Pop Quiz

4.3.2 There was Increased Confidence with Understanding of Learned Concepts. Prior to evaluating changes in reported confidence levels of participants, reliability (Cronbach Alpha) was calculated for the six items measured to ensure they were all consistent proxies for confidence. This reliability analysis showed consistently high reliability across all three time periods (pre-test = 0.910, post-test = 0.942, follow-up = 0.820). As such, the six items were averaged into a single overall measure of confidence and used in paired sample t-tests in a similar fashion to performance scores above. Participants' confidence with understanding of the concepts significantly increased by 48 percent, from an average confidence of 4.21 (on a 7-point scale) to 6.21 by the end of the activity (ΔConfidence = +2.00, two-tailed p-value = 0.000, n = 23, Cohen's d = 1.45). However, there was a marginally significant decline in confidence between the post-test and the follow-up evaluation (ΔConfidence = -0.43, two-tailed p-value = 0.083, n = 15, Cohen's d = 0.38). This finding is discussed in a later section.

4.3.3 High Levels of Enjoyment Reported, but also some Confusion. When participants (n = 23) were asked during the post-test what their perception about the exercise was like, 91 percent agreed or strongly agreed that they learned a lot from the card exercise. Two respondents (9%) only agreed 'somewhat.' Similar numbers held for the statement "I found the card exercise quite enjoyable." When participants were asked if they found the card exercise confusing, the results were more mixed. Twenty two percent of participants agreed or strongly agreed that the exercise was confusing, while only 39 percent disagreed or strongly disagreed. Thirty nine percent were split between the middle (somewhat agree, neither agree nor disagree, and somewhat disagree). This finding is also discussed in more detail in a later section.

4.3.4 High Level of Unaided Recall, Strong Levels of General Recall. To evaluate unaided recall, the open-ended written statements of participants were reviewed. While some participants were more verbose than others, there was a clear indication that important concepts had been learned. For instance, one participant who recalled the early part of the exercise focusing on changing business requirements wrote: "…using the playing cards allowed us to see that as needs change the database and how we retrieve information from it need to be flexible. This is due to the nature of business, which is always changing and growing." Another participant, who recollected the distributed processing aspects of the exercise, wrote: "Yes, I remember the explanation of increasing computational power with one machine vs spreading out computational power across many machines to have the same result." Fourteen out of the 15 follow-up responses demonstrated unaided recall of some concepts from the ELA. Lastly, the guided recall questions were also evaluated to confirm they were internally consistent (Cronbach alpha = 0.944), and their average was computed. Seventy three percent of participants had an average evaluation of somewhat agree, agree, or strongly agree to statements such as: "My memories from participating in 'the data shuffle' seem stronger than they might have been from simply reading or listening alone" and "I believe 'the data shuffle' sparked additional interest in this area." Only 13 percent somewhat disagreed or strongly disagreed with those statements.

5. DISCUSSION & CONCLUSION

In the preceding sections, an ELA using a regular deck of playing cards to practically illustrate business and technical concepts pertaining to data management is described. This approach takes abstract and difficult to explain concepts and makes them come to life in the form of an experiential learning activity (ELA) that is suitable for technical and non-technical students alike.

The authors' personal anecdotes and experiences from repeatedly carrying out this exercise reveal greater engagement, excitement, and involvement than the traditional speaking format that is the default way of teaching such concepts in both academic and professional settings. Even participants with prior, basic knowledge about data management in business report finding the use of tangible playing cards to be helpful. Three iterations of a survey were conducted (pre-test, post-test, and follow up survey after one week). Overall, the findings show that the ELA was effective. After completing the ELA, participants performed significantly better on a pop quiz covering data management concepts. When the quiz was repeated about a week later, during the follow-up survey, there was no significant change in performance. This means the insights from participating in the ELA were both immediate and memorable enough to sustain performance a week later without any significant drop. There was also an immediate boost in confidence reported by participants between the pre-test and post-test. Participating in the ELA boosted their beliefs that they could take an exam on the concepts learned. Importantly as well, participants generally enjoyed the exercise and were able to recall new concepts unaided. They also self-reported high levels of broad recall of the concepts learned.

However, there are also limitations to be aware of when conducting this ELA. First, the surveys showed that about one in five participants found the entire exercise somewhat confusing. From the author(s) experience conducting this ELA, the visual aids/supporting slides serve to reduce this confusion. Also, it is helpful to repeat the instructions multiple times at each phase of the exercise while walking around the class. Such practical steps will help reduce confusion among participants while facilitating learning and enjoyment. Second, it is important to note that there was a marginally significant reduction in confidence when the follow-up survey was conducted after a week ($p < 0.10$, two-tailed). It is unlikely that this ELA alone can build sustained confidence and deep insights into all important data management concepts, and therefore this exercise should be part of a broader learning program that can lead to sustained high levels of confidence with the material. For instance, Appendix C contains examples of existing courses that focus on concepts covered in this ELA. This described ELA can be used as a fun ice breaker to kick off such courses and build up interest ahead of detailed treatment of these concepts. Thirdly, the authors only have experience conducting this ELA in small to medium groups (around 40 maximum) with only one facilitator. Conducting this with a larger group (such as a large introductory class) may require additional assistants. Future work applying this specific method and other physical artifacts (e.g., Jenga blocks and Lego bricks) to teach other important concepts to both novices and experienced audiences of different sizes is strongly encouraged. Lastly, future work may consider evaluating different modes of instruction in comparison to each other. Due to the nature of the participants in this study (a university course), an effective control group or group with an alternative mode of instruction was not included in evaluation of effectiveness.

Beyond being a guided tutorial for those seeking to teach the specific concepts described above, this exercise should be regarded as part of broader efforts to use engaging ELAs to teach data concepts, both for business and technical audiences. With some thought and creativity, this exercise can be extended to teach other new technology concepts as well. The modular nature of the ELA described makes it possible to select only certain parts of the exercise and then combine newly developed routines. Further, this is a demonstration of the value of reverse-skeuomorphism in teaching digital concepts in a physical and tangible way.

6. REFERENCES

Ammar, S. & Wright, R. (1999). Experiential Learning Activities in Operations Management. *International Transactions in Operational Research*, 6(2), 183–197.

Anderson, J. (2013a). *Learn MapReduce with Playing Cards*. Retrieved from https://youtu.be/bcjSe0xCHbE.

Anderson, J. (2013b). *Processing Big Data with MapReduce*. Retrieved from https://pragprog.com/screencasts/v-jamapr/processing-big-data-with-mapreduce.

Baker, A., Navarro, E. O., & Van Der Hoek, A. (2003). An Experimental Card Game for Teaching Software Engineering. In *Proceedings of the 16th Conference on Software Engineering Education and Training, 2003. (CSEE&T 2003),* pp. 216–223, IEEE.

Baker, A., Navarro, E. O., & Van Der Hoek, A. (2005). An Experimental Card Game for Teaching Software Engineering Processes. *Journal of Systems and Software*, 75(1), 3–16.

Connolly, T. M. & Begg, C. E. (2007). Teaching Database Analysis and Design in a Web-Based Constructivist Learning Environment. In Filipe, J., Cordeiro, J., and Pedrosa, V. (eds.) *Web Information Systems and Technologies*, pp. 343–354. Berlin: Springer.

Connolly, T. M., Stansfield, M., & McLellan, E. (2006). Using an Online Games-Based Learning Approach to Teach Database Design Concepts. *Electronic Journal of E-Learning*, 4(1), 103–110.

Cook, L. S. & Olson, J. R. (2006). The sky's the limit: An activity for teaching project management. *Journal of Management Education*, 30(3), 404–420.

Davis, B. G. (2009). *Tools for Teaching*. Hoboken, NJ: John Wiley & Sons.

Dean, J. & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *6th Symposium on Operating Systems Design & Implementation*, San Francisco, CA.

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

Enders, C. K., Stuetzle, R., & Laurenceau, J.-P. (2006). Teaching Random Assignment: A Classroom Demonstration using a Deck of Playing Cards. *Teaching of Psychology*, 33(4), 239–242.

Gartner, W. B. & Naughton, M. J. (1988). The Deming Theory of Management. *Academy of Management Review*, 13(1), 138–142.

Gaskin, J. & Berente, N. (2011). Video Game Design in the MBA Curriculum: An Experiential Learning Approach for Teaching Design Thinking. *Communications of the Association for Information Systems*, 29(1).

Guimaraes, M. (2006). The Kennesaw Database Courseware (KDC): Strong Points, Weak Points, and Experience using it in a Classroom Environment. *Journal of Computing Sciences in Colleges*, 21(3), 91–96.

Hakulinen, L. (2011). Card Games for Teaching Data Structures and Algorithms. In *Proceedings of the 11th Koli Calling International Conference on Computing Education Research,* pp. 120–121, ACM.

Hamey, L. G. (2003). Teaching Secure Communication Protocols using a Game Representation. In *Proceedings of the fifth Australasian Conference on Computing Education-Volume 20*, pp. 187–196, Australian Computer Society, Inc.

Harrison, G. (2015). *Next Generation Databases: NoSQLand Big Data*. New York, NY: Apress.

Heim, G. R., Meile, L., Tease, J., Glass, J., Laher, S., Rowan, J., & Comerford, K. (2005). Experiential Learning in a Management Information Systems Course: Simulating IT Consulting and CRM System Procurement. *Communications of the Association for Information Systems*, 15(1).
Huang, C. D. & Behara, R. S. (2007). Outcome-Driven Experiential Learning with Web 2.0. *Journal of Information Systems Education*, 18(3), 329-336.
Kernochan, W. (2011). Why Most Business Intelligence Projects Fail, *Enterprise Apps Today*. Retrieved Nov 22, 2017, from http://www.enterpriseappstoday.com/business-intelligence/why-most-business-intelligence-projects-fail-1.html.
Kolb, D. (1984). *Experiential Learning as the Science of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.
Murray, M. & Guimaraes, M. (2008). Animated Database Courseware: Using Animations to Extend Conceptual Understanding of Database Concepts. *Journal of Computing Sciences in Colleges*, 24(2), 144–150.
Schibrowsky, J. A. & Peltier, J. W. (1995). The Dark Side of Experiential Learning Activities. *Journal of Marketing Education*, 17(1), 13–24.
Scott, S. (2012). Reverse-Skeuomorphism: When "Digital Things" Transcend the Screen to Analogue Devices. Retrieved August 24, 2018, from http://www.gizmodo.co.uk/2012/10/reverse-skeumorphism-when-digital-things-transcend-the-screen-to-analogue-devices/.
Turner, R. (1998). The Red Bead Experiment for Educators. *Quality Progress*, 31(6), 69.
Waite, D. (2011). A Simple Card Trick: Teaching Qualitative Data Analysis using a Deck of Playing Cards. *Qualitative Inquiry*, 17(10), 982–985.

AUTHOR BIOGRAPHIES

David Agogo is an assistant professor in the Information

Systems and Business Analytics Department at the Florida International University College of Business. His research focuses on studying unintended consequences of technology use and applying data science methods to analyze new and interesting research questions for practice and policy. He teaches courses in project management and business analytics at FIU. His work has appeared in *European Journal of Information Systems*, *Health Marketing Quarterly* and proceedings of *Association of Information Systems* and *Association for the Advancement of Artificial Intelligence* conferences/workshops.

Jesse Anderson Jesse Anderson is a Data Engineer, Creative

Engineer, and Managing Director of Big Data Institute. He works with companies ranging from startups to Fortune 100 companies on Big Data. This includes training on cutting edge technologies like Apache Kafka, Apache Hadoop, and Apache Spark. He has taught over 30,000 people the skills to become data engineers. He is widely regarded as an expert in the field and for his novel teaching practices. Jesse is published on O'Reilly and Pragmatic Programmers. He has been covered in prestigious publications such as *The Wall Street Journal*, *CNN*, *BBC*, *NPR*, *Engadget*, and *Wired*.

APPENDIX A: Screenshots of PowerPoint Slides

Figure A.1 shows the instructions that are provided centrally to the class, and the key points of the stage of the exercise are animated in after the discussion has concluded. Similarly, the use of prompt cards (Figure A.2) are a possibility. We have also included a link to a Kahoot.com quiz-show created instead of PPT slides: https://play.kahoot.it/#/k/4fe13692-b96a-47bb-aaa8-7ad1d4980714.

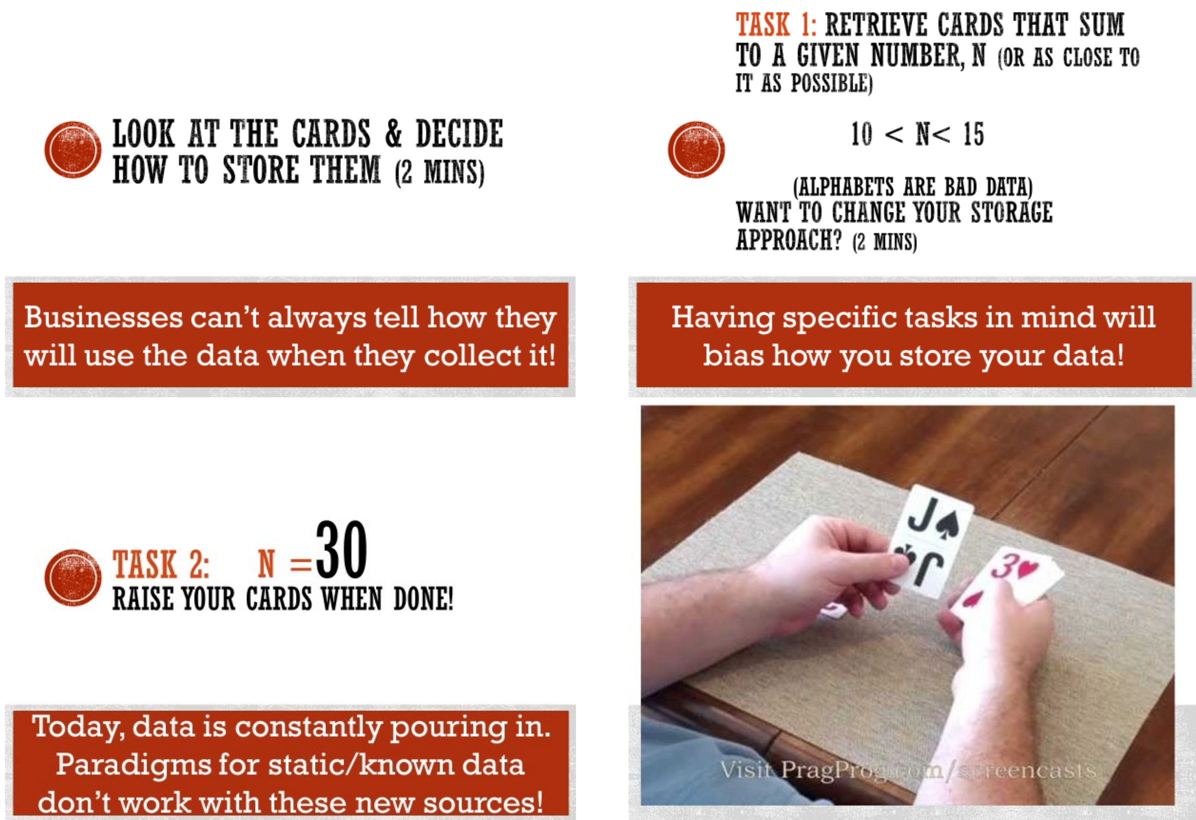

Figure A.1: Screenshots of Slides

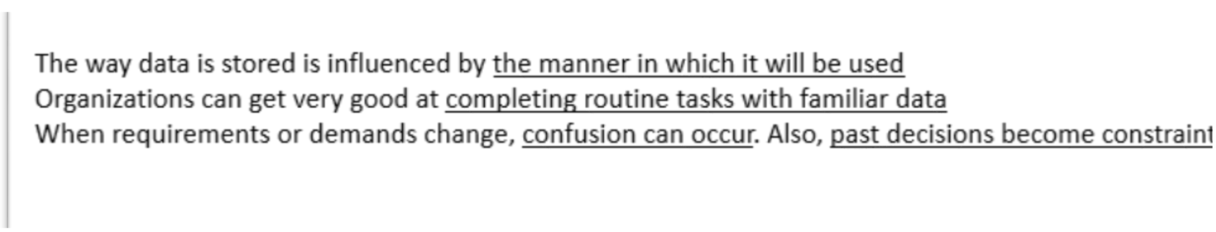

Figure A.2: Prompt Cards

APPENDIX B: Follow-Up Survey

B1. Confidence: If given an exam today, I would receive full credit for the question listed… (Strongly Disagree → Strongly Agree, 7 point)

- Why is data management important to a business?
- What is distributed processing?
- What is distributed storage
- What is MapReduce?
- How is static data different from streaming data?
- Describe what an algorithm is?

B2. Pop Quiz (Test of Understanding)

- It is fairly easy for organizations to anticipate how their data assets will be used ahead of time (T/F)
- It is always best to decide how to store and analyze data before collecting it (T/F)
- Current data practices of a business have a limited impact on their ability to deal with future developments (T/F)
- Systems that are used for static data needs can be easily repurposed to deal with streaming data (T/F)
- The most practical way to scale systems to deal with greater workloads is to get larger, more expensive equipment and software (T/F)

Which of the following is a unique challenge with distributed storage of information?

- I don't know what distributed storage means
- Coordinating versions of documents across different data partitions
- The cost of having multiple servers
- The time delay of copying data over the network

Which of the following is a challenge with distributed processing of information?

- I don't know what distributed processing means
- Hand off of results and communication across different points of processing
- The cost of having multiple processors working on the same task
- The time delay of waiting for all processors to complete working on an operation

Which of the following best describes an algorithm?

- I don't know what an algorithm is
- A list of steps that can be carried out repeatedly to complete a task
- A complex computer operation configured by computer scientists and programmers
- The output of a database CRUD (create, read, update or delete) operation

Attitudes towards exercise: How well do the following statements describe your feelings towards the card exercise from today's class? (Strongly Disagree → Strongly Agree, 7 point)

- I learned a lot from the card exercise
- I found the card exercise quite enjoyable
- I found the card exercise to be confusing

Unaided Recall: Can you mention any lessons/insights you can remember from 'the data shuffle' exercise carried out on the first day of class?

Recall: How well do the following statements apply to you? (Strongly Disagree → Strongly Agree, 7 point)

- I believe I can remember a lot about 'the data shuffle' exercise from the first day of class
- My memories from participating in 'the data shuffle' seem stronger than they might have been from simply reading or listening alone
- I believe 'the data shuffle' sparked additional interest in this area
- If given the chance, I'd like to repeat 'the data shuffle again'
- I can probably explain some important concepts learned from 'the data shuffle' in my own words

General Comments

Do you have any comments about "The Data Shuffle"? For instance, what did you particularly like about the exercise, what did you not like about it. An honest answer will help the professor improve this exercise for future classes.

APPENDIX C: Examples of Existing Courses that Could Integrate this Exercise

1. **University & Course:** Florida International University, Miami - ISM 4314 – Database Systems and Physical Design
   Week One: Business and Data
   Reference: Authors

2. **University & Course:** UNLV IS 471/MIS 671 – Big Data
   Week One: The relevance of big data; why big data
   Week Seven: Hadoop
   Reference: https://www.unlv.edu/sites/default/files/degrees/syllabi/RepresentativeSyllabi-MS-DataAnalyticsAppliedEconomics-MIS671.pdf

3. **University & Course:** IBM Big Data University -Big Data 101
   Module Two: Characteristics of big data – The Four Vs, understanding Big data with examples,
   Module Five: Text Analytics and Data Streams
   Reference: https://cognitiveclass.ai/courses/introduction-to-big-data/

4. **University & Course:** Stevens Institute of Technology BIA 676
   Week Three: Stream data model & examples
   Reference:
   https://www.stevens.edu/sites/stevens_edu/files/BIA%20676%20Data%20Stream%20Analytics%20Fall%202014.doc

5. **University & Course:** MIT Professional Course: Tackling the Challenges of Big Data
   Week Three: Distributed computing platforms
   Week Five: Algorithms for very large datasets and streaming computation
   Reference:
   https://mitxpro.mit.edu/c4x/MITProfessionalX/6.BD_2X/asset/Syllabus_for_Tackling_the_Challenges_of_Big_Data__February_3_-_March_17__2015_.pdf

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