Received: 6 April 2018 | Revised: 14 February 2019 | Accepted: 20 February 2019

DOI: 10.1002/acp.3539

RESEARCH ARTICLE

WILEY

Judging credibility: Can spaced lessons help students think more critically online?

Vanessa Foot-Seymour[1,2,3] | June Foot[3] | Melody Wiseheart[1,2]

[1] Department of Psychology, York University, Toronto, Canada

[2] LaMarsh Centre for Child and Youth Research, York University, Toronto, Canada

[3] York Region District School Board, Aurora, Canada

Correspondence
Melody Wiseheart, Department of Psychology, York University, 4700 Keele Street, Toronto, ON M3J1P3, Canada.
Email: melodywiseheart@gmail.com

Summary

Despite the prevalence of the spacing effect in the psychological literature, the impact of lesson timing has not yet been fully explored in real classrooms. The current study examined whether spacing could improve long-term retention of both factual and critical thinking curriculum-based teaching materials for children. Students 9 to 12 years old were taught to judge the credibility of websites in either three consecutive days of lessons or one lesson per week. Thirty-five days after the final lesson, students were tested on factual knowledge and applied their knowledge to evaluating a new website. Students in the spaced condition remembered more facts from the lessons and were better able to explain their website ratings than students in the massed group.

KEYWORDS

classroom, credibility, critical thinking, distributed practice, higher-level thinking, spacing effect

1 | SIGNIFICANCE STATEMENT

In the education system, student success is heavily dependent on foundational knowledge of the subject matter. Students are expected to retain taught content for later quizzes, assignments, end-of-unit tests, and in subsequent grades, during learning of new content that builds on existing knowledge. Because retention of information is crucial for learning, forgetting can be a problem for students and teachers when information needs to be retaught before moving forward in the curriculum. It is essential that educators are equipped with strategies to teach in a manner that efficiently maximizes learning and reduces forgetting. The psychology literature offers spacing of learning episodes (also known as distributed practice) as a potential solution to the issue of student forgetting. Spaced lessons improve long-term memory of the material by slowing down forgetting. This study investigated whether spacing of learning sessions could help students' long-term retention of both facts and critical thinking skills (i.e., website evaluation). This study tested the generalizability of existing spacing theories with children and using curriculum-based materials, allowing for practical recommendations for teachers, educators, and policy makers.

2 | INTRODUCTION

2.1 | Spacing effect

The spacing effect is a long-term memory advantage that occurs when information is relearned or reviewed in small, distributed blocks of time, as opposed to all at once (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). It is one of the oldest and best-documented phenomena in the history of learning and memory research (Cepeda et al., 2006; Wiseheart et al., 2019). A real-world example of spacing is seen when deciding whether to cram right before a test or exam or spread out learning. If a student has a designated amount of time to study, long-term retention would be better if the student were to space out review in equal intervals than to

spend the whole time studying in a more condensed block (Kapler, Weston, & Wiseheart, 2015).

In traditional spacing effect studies, a set of to-be-learned information is learned and then relearned in two sessions separated by a manipulated period of time known as the interstudy interval. After another period of time known as the retention interval, a final test is administered to assess level of retention (Figure 1).

There are several theories that explain the spacing effect. Glenberg (1979) believed that each item is stored in memory along with the specific context that it was learned in and that context changes over time. Glenberg's encoding variability theory supposes that the greater the number of unique contexts that are associated with each item, the larger the probability that item can be accessed in order for the information to be retrieved. Alternatively, study-phase retrieval theory (Thios & D'Agostino, 1976) suggests that learning of an item will be greater if the first memory trace can be retrieved from memory and that initial memory trace strengthened. For items that are retrieved soon after the first learning session, the reconstruction process will be easy, leading to little additional memory trace strengthening. For items that are retrieved later, after a spacing interval, retrieval will be more effortful, and greater reconstruction will occur. Some recent theories of the spacing effect combine encoding variability and study-phase retrieval accounts (e.g., Delaney, Verkoeijen, & Spirgel, 2010; see also Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009).

2.2 | Spacing effect in the classroom

In Ontario, Canada, the kindergarten to Grade 12 school year extends over a 10-month period, during which every teacher is asked to follow general and specific curriculum targets for student learning. Teachers are required to cover a significant amount of material in a condensed period of time, and they are given freedom to choose when and how to implement the expectations. Because of this, it is of great importance that they are given appropriate strategies to maximize learning and retention. Otherwise, increased time will be spent reteaching curriculum content that students have forgotten.

Although the spacing effect is a robust phenomenon, there is not yet enough research on how and whether it will work as expected when implemented in "regular" classrooms (Dempster, 1988). Some of the spacing effect studies that have been conducted in the classroom thus far have shown that the spacing effect is robust from preschool to university, across a wide variety of content areas. Spacing effects have been shown for vocabulary (Bloom & Shuell, 1981; Sobel, Cepeda, & Kapler, 2011), spelling (Fishman, Keller, & Atkinson, 1968), word lists (Zechmeister & Shaughnessy, 1980), addition (Reed, 1924), multiplication (Rea & Modigliani, 1985), and geometry (Rohrer, 2009; Rohrer & Taylor, 2007; Yazdani & Zebrowski, 2006). The spacing effect has achieved so much confidence from the psychological community that in 2016, the (U.S.) National Council on Teacher Quality (NCTQ; Greenberg, Pomerance, & Walsh, 2016) published a list of six core strategies that every teacher should be implementing in their classroom, and two of them were associated with spacing: repeatedly alternating solved and unsolved problems (which is referred to as interleaving, a type of spacing) and distributed practice (a synonym for spacing).

2.3 | Spacing effect: Beyond fact learning to critical thinking

It is important to highlight the difference between the spacing studies listed above, most of which taught students factual material (Cepeda et al., 2006) and those that have asked students to think critically. Fact learning is only one aspect of the learning process—students also need to be able to apply what they know to real-world situations that require independent thinking in problem solving situations. Students need to be able to explain, evaluate, analyze, and consider alternative perspectives and be confident in doing so. Before we recommend to teachers that they implement spacing in the classroom, it is necessary to demonstrate efficacy with major domains, including critical thinking skills. It is not a foregone conclusion that spacing will affect all domains, as some do not benefit from spacing (e.g., emotional skills; Wiseheart et al., 2019). This study will explore both fact learning and critical thinking, in order to gain deeper insight into the spacing effect.

A few studies have explored spacing in the classroom using curriculum-based materials that extend past simple fact learning; only two of these studies involved children. However, spacing research needs to be done methodically in order to properly test its effectiveness—the number of sessions, interstudy intervals, and retention intervals need to be chosen very carefully. Researchers must make wise choices about the combination of interstudy intervals and retention interval in order for a spacing effect to occur, as spacing benefits increase as interstudy interval increases to a maximal point of benefit, and then performance becomes worse when interstudy interval becomes even longer than the point of maximal benefit (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008).

Two studies that looked at spacing beyond fact learning in children were conducted by Vlach and Sandhofer (2012) and Gluckman, Vlach, and Sandhofer (2014). These studies had two goals: The first was to examine the effects of spacing, forgetting, and generalization across spacing intervals; the second was to determine whether spacing could promote simple and complex generalizations of concepts (Vlach, 2014). In the first study by Vlach and Sandhofer (Vlach et al., 2012),

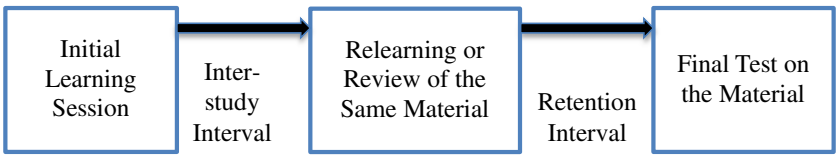

FIGURE 1 Visual representation of a traditional spacing study design [Colour figure can be viewed at wileyonlinelibrary.com]

children aged 5–7 years were taught to memorize facts about food chains in either massed learning sessions (all lessons were presented in 1 day), clumped (lessons were divided across 2 days), or spaced (lessons were divided across 4 days). Children were tested after 1 week had passed. Relative to students in the massed condition, students in both the clumped and spaced conditions improved in making both simple and complex generalizations.

In a replication by Gluckman et al. (2014), children aged 5–7 years were taught similar food-chain-related materials from the science curriculum to see whether spacing could help them to make generalizations to novel exemplars of a category. Students were taught in either massed, clumped, or spaced learning sessions. This study extended the 2012 study by adding a memory component that looked at basic retention. After the 1-week retention interval, students completed forced-choice questions, cued-recall questions, and free-recall questions based on the material they had learned in the lessons. Their memory for facts was significantly enhanced in all conditions (massed, spaced, and clumped), and as expected, there was a clumping and spacing advantage for both simple and complex generalizations. Massing did not show the same benefits.

In another study, Kapler et al. (2015) invited participants to attend a simulated university lecture. In this lecture, students learned both factual and higher-order thinking material and reviewed it after either a massed or spaced schedule. Results demonstrated that reviewing the material with more time in between the review sessions led to better long-term retention for both facts and higher-order thinking tasks. This study used higher-order thinking curriculum materials, relevant interstudy intervals (1 day and 1 week), a long retention interval (1 month), and time-efficient review methodology (an online quiz). The primary limitation of this study was that it took place in a mock classroom and thus utilized more experimental control than would be expected in the real world. However, it provides clues on how spacing works with higher order materials.

2.4 | Teaching critical thinking

Teaching critical thinking is often a challenge for educators, because it can be a difficult to break the concept down in a way that makes it applicable for classroom use. A qualitative study by Descours (2013) found that teachers use varied definitions of critical thinking. Teachers who were surveyed agreed that thinking critically is a skill, that it can be taught, and that it should be infused within the curriculum, but they had conflicting ideas on how to achieve these goals. Group work, class discussion, use of open-ended questions, and willingness to accept multiple perspectives were some of the most common suggestions.

Though there are many definitions of critical thinking, modern use of the term in educational contexts emerges from the work of philosopher John Dewey in the early 20th century. In his book, *How We Think*, Dewey (1910) described what it means to think critically. A good critical thinker, according to Dewey, carefully listens to the beliefs and opinions of others and takes time to thoroughly investigate all possible aspects before deciding whether or not they agree with their perspective. The necessity for good critical thinkers has not changed since then, and many scholars have discussed the importance of teaching critical thinking in schools.

Similarly, Ennis (1987) suggests that critical thinking is a general skill that involves "reasonable, reflective thinking that is focused on what to believe or do" (p. 10). Critical thinking, according to Ennis, involves a set of skills and dispositions that should be taught explicitly and then infused into everyday life to create an implicit understanding. Ennis refers to critical thinking as purposeful, self-regulatory judgment resulting from making an evaluation of an item. We built our definition of critical thinking in the present study upon Ennis' (1987) framework.

2.5 | Teaching critical thinking using website evaluation

Abrami et al. (2008) conducted a meta-analysis on the effects of instructional interventions on students' critical thinking skills. They looked at 177 studies and found that instruction of critical thinking was most effective when students were taught critical thinking instruction and subject content in approximately equal parts. This finding suggests that teachers should integrate critical thinking lessons with specific course content, so that students are able to put critical thinking skills into context and use them before transferring them to other disciplines. In addition, students should be given practical and relevant examples of when they might use their developing critical thinking skills. In line with this approach are the findings from Facione (1990), Halpern (1998), and Paul (1992).

In order to meet the goal of giving young students a content-specific and practical way to use their critical thinking skills, we decided that students would evaluate the credibility of websites. Determining website credibility is a relevant topic for students, because the nature of education has changed since the advent of widespread Internet usage. Paul (1992) suggested that the world is in ever-accelerating change; information is multiplying as it swiftly becomes obsolete and out of date. In the past decade, technology has advanced to become a regular part of student life. A Pearson study polled over 2,300 students in the United States 8 to 18 years old and determined that 99% of students use digital technology for school purposes, including desktop computers, laptops, netbooks, tablets, smart phones, and e-readers (Harris Poll, 2014). Out of these students, 70% used their devices for conducting research on the Internet. This poll demonstrates that the majority of students have virtually unlimited access to information on the Internet both at school and at home. As convenient as this access may seem, the ability to locate information can be problematic if students are not able to make informed decisions about whether that information is trustworthy or not. As such, there is a constant and growing need for students to obtain a more critical eye towards website content instead of simply accepting the thoughts and opinions that they are exposed to. Additionally, the Ontario Ministry of Education (2005) discusses the importance of training students in critical literacy. They advocate

that the "impact and influence of mass media and popular culture and the messages they convey, both overt and implied, can have a significant impact on students' lives. For this reason, critical thinking assumes a special significance" (p. 13).

2.6 | Current study

This study explored whether the spacing effect is robust enough to transfer to real-world classrooms using both fact learning material and critical thinking skills. Students 9 to 12 years old were taught to evaluate and judge the credibility of websites, and students practiced these skills with new websites during subsequent sessions. The lessons took three blocks of time that were scheduled on either three consecutive days of the week (massed) or 3 weeks in a row (spaced; 1 day per week). In each condition, students explored several different websites, were guided on how to evaluate and judge website credibility, and were taught how to locate evidence from the websites that helped them defend their decisions. Specificity in the lessons came from Ennis' (1987) framework, which defines judging credibility as the ability to think critically and make a decision about a source by asking questions about points of view, conflicts of interest, scientific information, methodologies, and assumptions. Ennis' framework provides a detailed list of pertinent dispositions and abilities, which have been implemented in validated educational assessment instruments (Ennis & Millman, 2005a, 2005b; Ennis, Millman, & Tomko, 2005; Ennis & Weir, 1985). Students were taught to identify four main categories of website evaluation (design, authority, content, and purpose) and were also given 17 more specific questions that would help lead them through a critical evaluation. Thirty-five days after the last lesson was completed, students were asked to evaluate a new website, so that we could see how many of the categories and questions they could remember (and use).

2.7 | Hypotheses

> **Hypothesis 1.** *The spacing effect will benefit critical thinking. Students in the spaced condition will use more information, by giving details taught in the lessons to explain their website ratings, than students in the massed condition.*

1a. At final test, students in the spaced condition will spontaneously use more of the four categories to explain their rating, compared with students in the massed condition. Students will demonstrate this knowledge in a paragraph.

1b. At final test, students in the spaced condition will spontaneously use more of the 17 questions to explain their rating, compared with students in the massed condition. Students will demonstrate this knowledge in a paragraph.

> **Hypothesis 2.** *The spacing effect will benefit fact learning. Students in the spaced condition will remember more information from the lessons than students in the massed condition.*

2a. At final test, students in the spaced condition will recall more of the four categories, compared with students in the massed condition. These students will be asked, "what are the four categories of website evaluation?" with four entry spots.

2b. At final test, students in the spaced condition will recognize more of the 17 questions, compared with students in the massed condition. Students will be given 34 questions (17 lures and 17 targets), and they will be asked to identify which questions they saw during the lessons. Hits-false alarms will be taken (a "hit" is a correct answer to a target item, and a "false alarm" is an incorrect answer to a lure) to prevent students from checking all 34 boxes and scoring 100%.

3 | METHOD

3.1 | Participants

Students 9 to 12 years old (Grades 4–6) were recruited from the York Region District School Board. This population was chosen because the Ontario curriculum (Ontario Ministry of Education, 2005) requires that at this point in their education, students must begin to "differentiate between fact and opinion; evaluate the credibility of sources, and recognize bias" (p. 89), but they have not yet had enough exposure to become proficient. Students were recruited from six schools across York Region. Participating schools were chosen based on principal and teacher interest. Overall, there were four Grade 4 classes, two 4/5 split classes (a mixture of Grade 4s and 5s in the class), three Grade 5 classes, four Grade 5/6 split classes, and seven Grade 6 classes. There were 93 Grade 4s, 110 Grade 5s, and 180 Grade 6s. Gender was approximately 50/50 male and female based on population demographics. Demographic census data demonstrate that 51% of York Region's population are Caucasian and 49% are from a visible minority (of those identifying as a visible minority, 45% self-identified as Chinese, 22% as South Asian, 8% as West Asian, 5% as Black, 5% as Filipino, 3% as Korean, 3% as Southeast Asian, 3% as Latin American, 2% as Arab, 1% as Japanese, and 4% as multiple or another visible minority). Additional details on York Region demographics are available on the Public Tableau website (Regional Municipality of York, 2018).

A total of 558 students within 20 classrooms participated in the lessons. Of the 558 students, four students did not receive parental consent for data to be used in research (resulting in a 99.3% consent rate). A total of 166 students were excluded from data analysis due to missing a lesson (e.g., due to illness) or being on an individualized education program that might affect results (e.g., needing to type responses on a computer). The final sample consisted of 388 students (n = 178, spaced; n = 210, massed). Our recruitment aim, which we nearly met, was n = 191 per group at analysis. We based sample size on an effect size of d = 0.37 and 95% power, our

estimate of the effect size for critical thinking, and spacing based on the most closely related prior classroom study (Kapler et al., 2015).

At the beginning of the lessons, students were given ID numbers and reminded that their participation was confidential. No names were written on any of the testing materials. Although lessons aligned with the Ontario curriculum, tests given during the study were not used as part of student grades, to avoid biasing course marks. Although marks were not used for student grades, all teachers reiterated that students were responsible for knowing these concepts and informed students that they could be tested on the concepts later as part of their coursework.

3.2 | Design

A between-subjects design was used, where classes were randomly assigned to either the spaced or massed condition, stratified as evenly as possible to ensure that classrooms in each condition were equal in ability. For example, if there were teaching partners (often referred to as *team teachers* because students are in the same grade, so materials are shared), these classrooms were separated and assigned to participate in each of the two conditions. Efforts were made to ensure a mixture of grades in each condition. For students requiring accommodations and modifications to teaching, adjustments were made in the classroom whenever necessary (extra help from volunteers or computer access if students needed to type instead of writing by hand), so that they could still participate in lessons. All participating students were taught as fairly and equitably as possible.

Students in both conditions were given an identical set of lessons, but with a timing manipulation. The massed condition received 3 days of lessons in a row, and the spaced condition received their lessons one per week for 3 weeks. Classes were taught any day of the week, depending on teacher and researcher availability. Day of the week was balanced across conditions to prevent confounds from day of week effects. Traditional spacing effect studies often have only one review session (for a comprehensive review of intervals, see Cepeda et al., 2006), but we added an additional review session so that students could experience more variety in the websites that they evaluated and to better implement curriculum guidelines. No strict time limits were imposed when students were evaluating websites, but all students completed the task within their language block (an average of 1.5 hr). We chose a 7-day fixed interstudy interval for practical reasons (i.e., it is feasible for teachers to implement a 1-week spacing design when lesson planning). Thirty-five days was used as the retention interval because it is optimal for a 7-day interstudy interval (Cepeda et al., 2008). Spacing effect benefits are present across a wide range of retention intervals, so many combinations of interstudy interval and retention interval would show essentially the same results (Cepeda et al., 2006, 2008).

All lessons were taught by an Ontario certified teacher (V.F.) with 10 years of classroom teaching experience. Lesson one included a class discussion followed by the pretest (Appendix A) and introduction to the checklist (which included room for a response paragraph at the end). In the discussion, students were taught the definition of credibility and brainstormed some examples of credible sources in their everyday lives. For the pretest, students were given a website and were asked to write how credible (i.e., trustworthy and believable) the website was and why. This was done to measure their responses at baseline. After the pretest, students learned about the checklist and were then asked to apply it to the pretest website and write their response about the same website's credibility, now that they knew what to look for. On the following 2 days, students practiced evaluating new websites using the same checklist. Thirty-five days after the last practice session, students took part in the final test. The final test had three sections: First, they were asked what the categories of website evaluation were, and we counted their correct answers out of four. Next, they were asked to look at a website and write a paragraph (without the checklist) to explain why they believed the website was credible or not. This was same worksheet as they saw during the pretest. Lastly, students saw a list of 34 questions and were asked to identify which 17 questions that they saw during the lessons.

3.3 | Materials

3.3.1 | Website evaluation worksheet

When designing a lesson plan, it is often suggested that teachers follow a backward planning design, meaning that all content should be taught with the final assessment in mind. As a result, every part of designing lessons for this study was structured in a way that stimulated learning leading up to the final test. At the end of the credibility lessons, students needed to know the following pieces of information:

1. Determining website credibility is not as easy as identifying whether sites are real or fake. Instead, evidence on the website helps students to make an informed decision that will most likely fall somewhere between 0 and 100.
2. Sorting evidence into four main categories can improve quality of evaluations.
3. Each of the four categories contains specific questions (17 in total) that can be treated as a website scavenger hunt (Table 1).

Based on these learning goals and a review of the website evaluation literature, a checklist was designed to help students learn how to generate a comprehensive rating of the website. The checklist included four categories and 17 questions (Table 1) and was based on a checklist by Bronstein (2007), which she created with the assistance of a Delphi panel of experts. Bronstein explored the validity and reliability of the checklist as a pedagogical tool, and she established the four main categories of website evaluation: design, authority, purpose, and content. The checklist was created in a way that would lead students to their final decision, guiding them along the way. The final decision about the website was a decision from 0 to 100, with 0 being not credible and 100 being credible (Appendix A). Bronstein recommended that website

TABLE 1 Four categories of website evaluation and 17 specific questions

Design
Do the pictures/photos and color choices look professional?
Are there any obvious spelling errors or typos?
Does the site appear to have what you were looking for?
Do the links to other pages and sites work properly?
Authority
Is the author or organization clearly marked?
Does the site tell you anything about the author? Does it tell you what their job is?
Do you believe that he/she is an expert?
Content
Does the site say when it was first created?
Does the site say when it was last updated?
Does the information match what you already know about the topic?
If you search the topic, can you find supporting evidence from another source?
Is this website appropriate for your grade level, or is it too difficult or mature for you?
Purpose
Has the author convinced you to see their point of view?
Do you think the author left out any important information?
Is the purpose of the site to teach you something new?
Is the purpose of the site to convince you to change your mind?
Is the purpose of the site to try to sell you something?

credibility checklists used for critical thinking purposes involve continuous scales rather than only yes/no answers, because critical evaluation is an ambiguous process that involves many different options for premises and different forms of reasoning that are equally legitimate. The goal is to gain deeper insight into students' thought processes. Neither the checklist nor the ratings from 0 to 100 were assessed or used in data analyses, because the critical thinking dependent variable in the study was the paragraph where they explained their answer. We hoped that students would use a combination of these tools (the checklist, rating scale, and written paragraph) in order to formulate an opinion.

3.3.2 | Websites

- Session 1: Introduction and pretest

Dog Island (Dog Island, 2013)

URL: https://web.archive.org/web/20150401070552/http://www.thedogisland.com:80/

From the website: Over 2,500 dogs are already enjoying a better life at Dog Island. Separated from the anxieties of urban life, dogs on Dog Island are healthy dogs who live a natural, healthy, and happy life, free from the stress and hardship associated with daily life among humans. They live with almost limitless space and tens of thousands of rabbits, rodents, fish, and other natural prey. Surrounded by thousands of other dogs, this is the only place for them to be truly social and create healthy families (Dog Island, 2013).

- Session 2: Relearning

Save the Pacific Northwest Tree Octopus (Zapato, 1998).

URL: https://web.archive.org/web/20150212012956/https://www.zapatopi.net/treeoctopus/

From the website: The Pacific Northwest tree octopus (*Octopus paxarbolis*) can be found in the temperate rainforests of the Olympic Peninsula on the west coast of North America. Their habitat lies on the eastern side of the Olympic mountain range, adjacent to Hood Canal. These solitary cephalopods reach an average size (measured from arm tip to mantle tip) of 30 to 33 cm. Unlike most other cephalopods, tree octopuses are amphibious, spending only their early life and the period of their mating season in their ancestral aquatic environment. Because of the moistness of the rainforests and specialized skin adaptations, they are able to keep from becoming desiccated for prolonged periods of time, but given the chance, they would prefer resting in pooled water (Zapato, 1998).

- Session 3: Relearning

Paleontologists Say Tyrannosaurs Were Cannibals (Prostak, Anderson, & Lazaro, 2011).

URL: https://web.archive.org/web/20151106222824/http://www.sci-news.com/paleontology/science-tyrannosaur-cannibalism-wyoming-03392.html

From the website: A distinctive pattern of tooth marks on a 66-million-year-old tyrannosaur bone found in eastern Wyoming offers one of the best evidences yet that tyrannosaurid dinosaurs were not shy about eating their own kind, according to a team of paleontologists led by Loma Linda University, California (Prostak et al., 2011).

- Final test

Mike the Headless Chicken (Cobb Strategic Marketing, 2013).

URL: https://web.archive.org/web/20150108110504/http://www.miketheheadlesschicken.org:80/

From the website: The amazing true story of this famous fowl dates back to September 10, 1945 when Mike, a young Wyandotte rooster, was about to become the dinner of Fruita, Colorado, farmer Lloyd Olsen. With a sharp axe in hand, Mr. Olsen firmly held Mike, preparing to make the bird ready for his wife Clara's cooking pot. Mr. Olsen swung the implement, thereby lopping off poor Mike's head. Mike shook off the event, then continued trying to peck for food. Mike's will to live remains an inspiration. It is a great comfort to know you can live a normal life, even after you have lost your mind (Cobb Strategic Marketing, 2013).

3.4 | Analyses

A set of *t* tests and Bayesian analyses were conducted. Before running analyses of interest, tests were conducted to ensure that assumptions were satisfied. The independence of observations assumption was violated because there was nesting by classroom, but this violation was addressed by running a separate post hoc analysis with students constituting repeated measures, looking at classrooms as independent data points. This post hoc analysis showed no difference from the primary analyses.

Further analyses were conducted on the data set due to violations of normality assumptions in the sample. A nonparametric test was used on the ranks (Mann–Whitney *U*) in order to ensure of the accuracy of the study results. These results showed the same outcome as the *t* tests. Data and materials are available at https://osf.io/jbyvz/.

4 | RESULTS

4.1 | Baseline

We examined use of the four categories and 17 questions in the Session 1 pretest rating explanation paragraph, in order to ensure that students did not already know the material and as a check on the sufficiency of random assignment and stratification procedures. Inter-rater reliability scores were measured with Pearson's *r*, and the degree of consensus was 0.81 for the four categories and 0.80 for the 17 questions. Students received a mark out of four for mentioning categories (paraphrasing was accepted). For example, if students said, "I saw who created the website," students got a mark in the authority category. Students were marked out of 17 on which specific questions they chose to include in their rating explanation. For example, students may have said, "I saw who created the website [adds their name] and I looked them up and it tells me who they are. I think they are experts on the topic and I believe what they are saying." This response would have received three marks for authority—one for mentioning the author's name, one for adding additional details about them, and one for deciding if the author is an expert. Massed and spaced groups did not differ on how many of the four categories were used during rating explanations, $t(375.15) = 0.23$, $p = 0.81$, $d = 0.026$, $BF_{10} = 0.116$ (moderate evidence for the null hypothesis), nor did groups differ in use of the 17 questions, $t(382.25) = 0.99$, $p = 0.32$, $d = 0.10$, $BF_{10} = 0.180$ (moderate evidence for the null hypotheses).

4.2 | Hypothesis 1

At final test, students were asked to rate the website from 0 to 100 and explain their rating in a paragraph. Students were marked on how many of the four categories and 17 questions that they referenced in their paragraph. One coder (V.F.) marked every paragraph, and a second rater marked them again for inter-rater reliability. There was a total of four other coders, one for every category (one coder looked for design, one for author, one for content, and one for purpose; each looked for the specific questions coinciding with the category they coded). Inter-rater reliability was calculated with Pearson's *r*, and the degree of consensus was 0.91 for the four categories and 0.90 for the 17 questions. Results showed that students in the spaced group used significantly more of the four categories than the massed group, $t(337.79) = 4.66$, $p < 0.001$, $d = 0.48$, $BF_{10} = 4{,}573.79$ (extreme evidence for a group difference), and the spaced group also used more of the 17 questions than the massed group, $t(359.62) = 4.74$, $p < 0.001$, $d = 0.48$, $BF_{10} = 5{,}365.80$ (extreme evidence for a group difference), to defend their ratings. See Table 2 for a summary. Appendix B contains examples of the paragraphs written by students, with comments. As previously mentioned, these paragraphs were marked by two raters to ensure test validity.

TABLE 2 Descriptives for pretest paragraphs, final test paragraphs, and final test fact learning measures

	Massed				Spaced			
	n	*M* (%)	*SD*	[95% CI]	*n*	*M* (%)	*SD*	[95% CI]
Pretest critical thinking								
4 categories[a]	208	10.5	15.3	[8.4, 12.6]	177	10.9	15.1	[8.6, 13.1]
17 questions[a]	208	5.9	6.1	[5.0, 6.7]	177	6.5	5.4	[5.7, 7.3]
Final test critical thinking								
4 categories[a]	210	17.2	21.6	[14.3, 20.1]	177	28.8	26.6	[24.9, 32.8]
17 questions[a]	210	18.6	11.0	[17.1, 20.1]	177	24.2	12.1	[22.4, 26.0]
Final test fact learning								
4 categories recalled[b]	186	30.8	28.8	[26.6, 34.9]	176	56.2	31.7	[51.5, 60.9]
17 questions recognized[c]	186	62.6	31.1	[58.1, 67.1]	177	68.4	27.2	[64.4, 72.5]

[a]Four categories and 17 questions spontaneously used in a paragraph, during both the pretest and final test, shown as a percentage.

[b]Cued recall asked students to "tell us what the four categories of website evaluation are."

[c]In a 34-question test, students had to recognize which of the 17 questions they saw during the lessons among 17 lures.

4.3 | Hypothesis 2

For the cued recall measure, we looked at the number of correct answers out of four. When students were asked to recall the four categories (design, authority, content, and purpose), students in the spaced group remembered significantly more than those in the massed group, $t(353.45) = 7.99$, $p < 0.001$, $d = 0.85$, $BF_{10} = \infty$ (extreme evidence for a group difference). In a recognition test that included the 17 questions and 17 lures (e.g., "Was the website slow to load?" which was not on the checklist), evidence was inconclusive whether percent of questions recognized differed between groups, $t(358.36) = 1.91$, $p = 0.057$, $d = 0.20$, $BF_{10} = 0.66$ (indeterminate whether a group difference was present).

5 | DISCUSSION

The current study explored the spacing effect in a classroom setting, using curriculum-based materials and course-relevant interstudy intervals and retention interval, taught by a certified teacher in a typical manner. All of the lessons were associated with critical thinking, based on the definitions described earlier. Immediately after students were shown the first website and learned that it was a hoax, their guards were up for the remainder of the lessons. Students became extremely critical of what they read online, and they searched for flaws in all aspects of the websites. Even though they may not always have been able to pinpoint the problems with what they encountered online, a critical thinking mindset was in place (according to the definition of critical thinking by Gilbert, 2014). Data were not collected to measure mindset; however, anecdotal comments by the class teacher (V.F.) note that students were extremely skeptical of even the smallest details.

In an effort to teach students how to identify key features of the websites, we taught them exactly what to look at and how to discuss what they saw with the purpose of coming to a specific outcome (i.e., completion of the rating, and explanation of the rating). This is closely linked to Ennis' (1987) perception of a good critical thinker—someone who uses reasonable, reflective thinking when deciding what to believe or do—and also Dewey (1910), who asked critical thinkers to be objective and consider their options while taking time to investigate before deciding whether or not they agree. In order to thoughtfully consider alternate options, students needed foundational materials that they could use to explain their perspective, which were the four categories and 17 questions that were taught during each lesson. Although students had to remember the questions and the categories, they also had to manipulate and apply these facts within their specific situation. There was a goal (evaluating the website and giving it a rating), and students used what they learned to meet this goal.

The outcome that students in the spaced condition used more of the categories and questions to explain their rating suggests that they were better able to communicate their perspective and explain why they chose the option (rating) that they chose. Appendix B shows example student answers from their website evaluations.

Because the spacing effect has been shown to be robust with fact learning material (Cepeda et al., 2006), we expected to see a benefit on fact learning measures. Our real-world classroom replication was successful, demonstrating that the spacing effect can generalize to a real-world classroom setting, and that both fact learning and critical thinking can benefit from spacing. However, only one of our two fact learning measures showed a spacing benefit. Our failure to find a benefit in the cued recognition measure may have been due to the large number of recognition questions listed on the final test (17 targets plus 17 lures) and the collection of recognition questions at the end of a 90-min class block, which could have resulted in student fatigue at time of testing.

At the beginning of the second and third review sessions, students were asked to recall the four categories and 17 questions of website evaluation through a group discussion. Although data were not collected during the discourse, anecdotal evidence suggested that students in the massed condition seemed to have a much more efficient discussion. This may have been because these students had been exposed to the information more recently. The spaced group, on the other hand, often struggled to remember the categories during the discussion (i.e., students needed lots of prompting and reminders). The discussions at the beginning of the lessons were vital because the relearning helped students to build a fundamental knowledge base, which students used through the remainder of the lessons and then on the final test after 35 days.

Our results are generally consistent with a real-world history classroom study. Carpenter, Pashler, and Cepeda (2009) found an effect size of $d = 0.49$ for fact learning, whereas the current study found an average effect size of $d = 0.58$ across the two fact measures. Kapler et al. (2015), a mock classroom study, found an effect size of $d = 0.33$ for fact learning of meteorology content. Thus, the corpus of classroom evidence suggests a robust spacing effect benefit in classroom learning, with a mean effect size of $d = 0.47$ across these three studies.

5.1 | Conclusion and future directions

This study addressed concerns that there is not yet enough research on the spacing effect in the classroom (Dempster, 1988) and that many previous classroom studies have not used educationally relevant inter-study intervals and retention intervals (Cepeda et al., 2006). In addition, this study looked at whether spacing effects can extend to include critical thinking skills. One of the primary goals of any education system is to help students remember the information they are taught at school so that they can apply it to their everyday lives. This may require students to hold on to specific pieces of information for extended periods of time, only to see it out of context and have to manipulate it to fit a certain situation. If classroom spacing effect studies support the enhancement of retention of these critical thinking skills, it might also help to reduce forgetting, which can sometimes cause barriers to student learning (Custers & ten Cate, 2011).

If spacing in the classroom proves useful, it is a relatively easy adjustment to make in classrooms. Teachers would not need to

directly adjust their lesson plans, but rather the timing of delivery to students. These course alterations could be done at the beginning of the school year while creating long-range plans, and only slight alterations need to be made in order to successfully implement spacing in the classroom.

However, this study was not without limitations. One limitation is the reality of conducting classroom research. Each classroom varies on many levels that are difficult to control, and student performance can depend on many factors. There are also occasional interruptions in the classroom. During this study, we had a fire drill, a lockdown procedure, three snow days that caused minor shifts session timing, and teachers were absent on certain days, which caused a break in student routine. These challenges may have affected results, but there is no way to determine to what extent. Because these types of interruptions are typical in many classrooms, and students are used to interruptions, this study accurately represented real-world implementation in a typical classroom setting.

A next step for classroom spacing effect research is to explore whether it can be effective on a wider scale, using different teachers and subjects. In this study, lessons were carried out by one Ontario certified teacher, to help control for teacher effects. In the real world, teachers use different teaching styles and have different personalities and approaches. Researchers need to investigate whether the results we found are observed when different teachers implement the same lesson plans, as would happen in the typical classroom. In addition, a variety of curriculum-based subject material should be used (inquiry math would be an excellent choice). If those factors can be explored, perhaps strong, evidence-supported recommendations can be made to start using spacing in the classroom. We hope that the present study will encourage future investigations on the spacing effect in real-world classroom learning.

AVAILABILITY OF DATA AND MATERIALS

The datasets supporting the conclusion of the article and study materials are available at https://osf.io/jbyvz/

AUTHOR CONTRIBUTION

All authors assisted in the design of the project. V.F. and J.F. designed the media literacy curriculum, and V.F. taught all of the lessons to students. All authors assisted in the creation of the assessment materials. V.F. and J.F. collected data with supervision from M.W. V.F. and M.W. wrote the manuscript. M.W. and V.F. analyzed data (M.W. ran primary analyses, and V.F. checked for assumptions and ran nonparametric test of ranks).

ACKNOWLEDGEMENTS

Thank you to members of the Cognitive Flexibility Lab for assisting with study implementation and data collection and to the York Region District School Board for allowing us to conduct research in the schools. Thank you also to each principal, teacher, parent, and student whose participation made this study possible.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ORCID

Melody Wiseheart https://orcid.org/0000-0001-8332-6775

REFERENCES

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M., Tamim, R., & Zhang, D. A. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage one meta-analysis. *Review of Educational Research*, *78*, 1102–1134. https://doi.org/10.3102/0034654308326084

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*, 245–248. https://doi.org/10.1080/00220671.1981.10885317

Bronstein, D. M. (2007). The efficacy of a web site evaluation checklist as a pedagogical approach for teaching students to critically evaluate internet content (Unpublished doctoral dissertation). The Graduate School of Computer and Information Sciences Nova Southeastern University.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*, 760–771. https://doi.org/10.1002/acp.1507

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*, 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Cobb Strategic Marketing (2013). *Mike the headless chicken: Home. In Mike the Headless Chicken*. Retrieved from https://web.archive.org/web/20151220235113/http://www.miketheheadlesschicken.org/

Custers, E. J. F. M., & ten Cate, O. T. J. (2011). Very long-term retention of basic science knowledge in doctors after graduation. *Medical Education*, *45*, 422–430. https://doi.org/10.1111/j.1365-2923.2010.03889.x

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of learning and motivation*, *53*, 63–147. https://doi.org/10.1016/S0079-7421(10)53003-2

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*, 627–634. https://doi.org/10.1037/0003-066X.43.8.627

Descours, K. (2013). 21st century pedagogy: A classroom perspective on critical thinking (Unpublished master's thesis). York University.

Dewey, J. (1910). *How we think*. New York: Prometheus Books.

Dog Island (2013). In *Dog Island Free Forever*. Retrieved from https://web.archive.org/web/20160315022057/http://www.thedogisland.com/

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). New York: W. H. Freeman and Company.

Ennis, R. H., & Millman, J. (2005a). *Cornell critical thinking test* (5th ed.). Pacific Grove, CA: Critical Thinking Books & Software.

Ennis, R. H., & Millman, J. (2005b)). *Cornell critical thinking test, level Z* (5th ed.). Pacific Grove, CA: Midwest Publications.

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell critical thinking tests: Administration manual* (Fifth ed.). Seaside, CA: The Critical Thinking Company.

Ennis, R. H., & Weir, E. (1985). *The Ennis–Weir critical thinking essay test*. Pacific Grove, CA: Midwest Publications.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. Newark, DE: American Philosophical Association. (ERIC Document Reproduction Service No. ED315423)

Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology*, *59*, 290–296. https://doi.org/10.1037/h0020055

Gilbert, M. A. (2014). *Arguing with people*. Peterborough, ON: Broadview Press.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112. https://doi.org/10.3758/BF03197590

Gluckman, M., Vlach, H., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology*, *28*(2). https://doi.org/10.1002/acp.2997

Greenberg, J., Pomerance, L., & Walsh, K. (2016). Learning about learning: What every new teacher needs to know (Rep.). Retrieved March 23, 2016, from National Council on Teacher Quality website: http://www.nctq.org/dmsView/Learning_About_Learning_Report

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist*, *53*, 449–455. https://doi.org/10.1037/0003-066X.53.4.449

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38–45. https://doi.org/10.1016/j.learninstruc.2014.11.001

Mozer, M. C., Pashler, H., Cepeda, N. J., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information systems 22*. San Diego, CA: Neural Information Processing Systems Foundation.

Ontario Ministry of Education. (2005). The Ontario curriculum grades 1–8: Language. [Program of Studies]. Retrieved January, 2016, from: http://www.edu.gov.on.ca/eng/curriculum/elementary/language18currb.pdf

Paul, R. W. (1992). Critical thinking: What, why, and how. *New Directions for Community Colleges*, (77), 3–24. https://doi.org/10.1002/cc.36819927703

Harris Poll (2014, May 9). Pearson student mobile device survey 2014: Grades 4 through 12 (Rep.). Retrieved March 25, 2016, from Pearson Education website: http://www.pearsoned.com/wp-content/uploads/Pearson-K12-Student-Mobile-Device-Survey-050914-PUBLIC-Report.pdf

Prostak, S., Anderson, N., & de Lazaro, E. (2011). Paleontologists say tyrannosaurs were cannibals. In *Paleontology*. Retrieved from https://web.archive.org/web/20160610042353/http://www.sci-news.com/paleontology/science-tyrannosaur-cannibalism-wyoming-03392.html

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning: Journal of Practical Research & Applications*, *4*, 11–18.

Reed, H. B. (1924). Repetition and association in learning. *The Pedagogical Seminary*, *31*(2), 147–155. https://doi.org/10.1080/08919402.1924.10532929

Regional Municipality of York (2018). 2016 census: Public tableau. Retrieved January 6, 2019, from https://public.tableau.com/profile/regional.municipality.of.york#!/vizhome/YorkRegionCensusProfile2016-Part1/Story1

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, 4–17.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498. https://doi.org/10.1007/s11251-007-9015-8

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*, 763–767. https://doi.org/10.1002/acp.1747

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behaviour*, *15*, 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/cdep.12079

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Wiseheart, M., Küpper-Tezel, C., Weston, T., Kim, A. S. N., Kapler, I. V., & Foot-Seymour, V. (2019). Enhancing the quality of student learning using distributed practice. In J. Dunlosky, & K. Rawson (Eds.), *Cambridge handbook of cognition and education* (pp. 550–584). New York: Cambridge University Press.

Yazdani, M. A., & Zebrowski, E. (2006). Spaced reinforcement: An effective approach to enhance the achievement in plane geometry. *Journal of Mathematical Sciences and Mathematics Education*, *1*, 37–43.

Zapato, L. (1998). The Pacific Northwest tree octopus. In *Help Save the Pacific Northwest Tree Octopus from Extinction!* Retrieved from https://web.archive.org/web/20160708102348/http://zapatopi.net/treeoctopus/

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*, 41–44. https://doi.org/10.3758/BF03329756

How to cite this article: Foot-Seymour V, Foot J, Wiseheart M. Judging credibility: Can spaced lessons help students think more critically online? *Appl Cognit Psychol*. 2019;33:1032–1043. https://doi.org/10.1002/acp.3539

APPENDIX A

ID:_____ Date: _____

Should I Trust This Website?
Critical Evaluation of Online Sources

You were asked to write a report about dogs. When you typed "nice places for dogs to hang out" into Google, you were brought to the website:

www.thedogisland.com

Have you ever seen or heard of this website before today? _____

On a scale of 0-100, how credible does the website look to you overall?

Your answer: ____ / 100

Please explain your rating using information from the website.

_____

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX B

EXAMPLES OF STUDENT ANSWERS

Tree Octopus
(Rating 30/100)
Please explain your rating using information from the website.
I gave 30% because the design was organized well and because it didn't say much about the author and his own webpage about himself is all comics and says nothing about him being an expert. It says when it was made and it was last updated in 2015 so the content seems reasonable. It's a good purpose to save the endangered animals but the things they want you to do are crazy! For example, "participate in Tree Octopus marches. You can demonstrate their plight during the march by having your friends dress up as tree octopuses while you attack them in a lumberjack costume."
Post: Mike the Headless Chicken
(Rating 15/100)
Please explain your rating using information from the website.
For authority it doesn't really show who made the website but it shows who designed the website. The purpose of the website well I don't know all the website is saying it what he did and how long he survived. For the links they work so that is good for the website. It also says that "mike is going for the gold" but that looks fake because he is dead now. It says that the designer is aha consulting so that shows that the design is real or probably real because I looked them up and they are a real thing. In the website it doesn't look like there is any spelling mistakes so that is good. The author looks like he is trying to persuade me that mike the headless chicken is real and he is also trying to persuade me to buy something like shirts and shows pictures and prices. To me I don't like the content and the pictures because I don't trust them and the fact that the author is showing me pictures of the headless chicken that all look fake. My last thing to me is that this website doesn't look credible.