Improving problem solving with retrieval-based learning

Garrett O’Day

Purdue University

Acknowledgements

I wish to thank Courtney Isgett, Cara Brookins, Alexandra Simpson, Caitlyn Gilbert, and Stacey Gall for their assistance in collecting and scoring responses. Courtney Isgett deserves additional praise for help adapting the materials and for her contributions to the development of the experimental protocol. I also thank Julia Zuchkov for her constructive feedback during the early drafts of this paper. My cat, Raja, also deserves recognition for the countless edits that she added without even being asked. I would also like to express my deep gratitude to Dr. Jeffrey Karpicke, Dr. James Nairne, and Dr. Thomas Redick for all of the wisdom they have imparted to me over my first three years of graduate school and for their guidance during the planning stages for this project.

Abstract

Recent research asserts that the mnemonic benefits gained from retrieval-based learning vanish for complex materials. Subsequently, it is recommended that students study worked examples when learning about complex, problem-centered tasks. The experiments that have evaluated the effectiveness of studying worked examples tend to overlook the mental processing that students engage in when completing retrieval-based learning activities. In contrast, theories of transfer-appropriate processing emphasize the importance of compatibility between the cognitive processing required by the test and the cognitive processing that is activated during learning. For learners to achieve optimal test performance, according to transfer-appropriate processing, they need to study in such a way that they are engaging in the same mental processing that will be required of them when tested. This idea was used to generate testable predictions that compete against the claim that the retrieval practice effect disappears for complex materials, and these competing predictions were evaluated in three experiments that required students to learn about the Poisson probability distribution.

In Experiment 1, students learned the general procedure for how to solve these problems by either repeatedly recalling the procedural steps or by simply studying them. The retrieval practice condition produced better memory for the procedure on an immediate test compared to the study only condition. In Experiment 2, students engaged in the same learning activities as Experiment 1, but the test focused on their problem-solving ability. Students who practiced retrieval of the procedural steps experienced no benefit on the problem-solving test relative to the study only condition. In Experiment 3, students learned to solve Poisson probability problems by studying four worked examples, by studying one worked example and solving three practice problems, or by studying one worked example and solving three practice problems with

feedback. Students were tested on their problem-solving ability one week later. The problem-solving learning activities outperformed the worked example condition on the final problem-solving test. Taken together, the results demonstrate a pronounced retrieval practice effect but only when the retrieval-based learning activities necessitated the same mental processing that was required during the final assessment, providing support for the transfer-appropriate processing account.

Introduction

Learning is often conceptualized as encoding or the acquisition of knowledge and can occur when students are reading textbooks, taking notes, or studying examples. Sometime after students have finished learning the material, they are given a test to measure what they have learned. Testing, in this view, is seen as nothing more than a diagnostic tool that students do not enjoy completing and teachers do not enjoy administering. Fortunately, research has consistently demonstrated that enhanced learning arises from taking tests and retrieving information, which vastly outweighs the drawbacks of testing. (for a recent review see Karpicke, 2017).

This counterintuitive finding, which students often fail to use to their advantage, is now commonly referred to as retrieval-based learning and has been found to improve test performance for students in elementary school (Marsh et al., 2012), middle school (McDaniel et al., 2013), high school (McDermott et al., 2014), college (Knouse et al., 2016), and medical school (Larsen et al., 2013). The retrieval practice effect has also been demonstrated across a wide range of materials and has been found to promote the learning of word pairs (Karpicke & Smith, 2012), key-term definitions (Lipko-Speeda et al., 2014), and educational texts (Blunt & Karpicke, 2014). Some studies have even examined retrieval-based learning with non-verbal materials such as Chinese characters with English speaking participants (Kang, 2010) and navigational routes (Kelly et al., 2015). Even though research continues to demonstrate the breadth of the retrieval practice effect, there is limited research investigating the efficacy of retrieval-based learning for problem-centered activities that are critical in STEM (Science, Technology, Science, and Mathematics) domains.

Learning by Studying Worked Examples

Alternative research has found that substituting traditional problem-solving tasks with the studying of worked examples, which consist of a problem statement along with a detailed explanation of how to solve the problem, improves future problem-solving performance (for reviews see Renkl, 2014; Renkl & Atkinson, 2010; van Gog & Rummel, 2010). This benefit of studying worked examples over solving practice problems is referred to as the worked example effect (Sweller, 2006; Sweller & Cooper, 1985). The primary theoretical explanation for the worked example effect comes from cognitive load theory, which assumes that humans have a limited working memory capacity that is vital for successful problem solving (Paas & van Gog, 2006). When solving practice problems, novice learners squander their limited cognitive resources because they lack the necessary understanding of problem-solving strategies (van Gog et al., 2011). They are, instead, forced to rely on shallow and ineffective strategies leading to a high mental burden referred to as cognitive load (Sweller, 1988). When learners are placed under cognitive load, they lack the cognitive resources that are critical for acquiring the necessary problem-solving strategies (Hanham et al., 2017). Worked examples, on the other hand, reduce this cognitive burden by providing learners with appropriate problem-solving strategies. By studying an expert solution, students do not need to waste their cognitive resources by searching for and applying a potentially erroneous strategy in an attempt to solve the problem. Instead, they can devote all of their attention and cognitive capacity to learning the worked examples and are better equipped to acquire the cognitive schema needed to successfully solve future problems (Sweller et al., 1998; van Gog & Rummel, 2010).

Van Gog, Kester, and Paas (2011) tested the benefit of worked examples for complex materials that required learners to troubleshoot electrical circuits. Students either studied four worked examples, solved four problems, alternated between two example-problem pairs, or alternated between two problem-example pairs. They found that the worked example only condition and the example-problem pairs condition not only decreased self-reported mental effort (i.e., cognitive load) but also improved test performance compared to the other conditions. These results were interpreted as evidence that studying worked examples is an effective learning activity compared to solving practice problems, at least on an immediate test.

Comparing and Contrasting the Worked Example Effect and The retrieval Practice Effect

Additional research on the worked example effect by van Gog and Kester (2012) noted the similarity between the example-problem pairs condition (Example-Problem-Example-Problem [EPEP]) in the experiment described previously and the traditional retrieval practice condition that requires students to alternate between studying and retrieving the material (Study-Test-Study-Test [STST]). The researchers argued that the example-problem condition is akin to a retrieval practice condition because students alternate between studying and solving problems, which requires learners to actively retrieve past information. However, a major difference between these two procedures is that traditional retrieval practice experiments provide students with an opportunity to review the exact information they were trying to remember after each retrieval opportunity (e.g., Roediger & Karpicke, 2006). In experiments that have used example-problem pairs, on the other hand, involve the presentation of a novel example after each practice problem. Studying a completely new example requires students to spend a portion of their limited time reading and understanding the demands for that specific question. Feedback for the exact practice problem would instead allow students to find their immediately preceding mistakes quicker as they would already be familiar with the problem, granting them more time to specifically hone their problem-solving knowledge.

Despite this major difference between the two procedures, the researchers predicted that the example-problem pairs condition would lead to better performance since it has opportunities for retrieval that are not present in the example-only condition. However, this prediction was not supported by the results from van Gog, Kester, and Paas (2011). In their study, the example-problem pairs condition produced equivalent test performance compared to the example only condition, which is not surprising because researchers have known for quite some time that repeated studying tends to be favored on immediate tests but is ultimately unsuccessful at cultivating long-term learning. In fact, one study (Roediger & Karpicke, 2006) asked learners to tally the number of times they read through a short passage during a retrieval practice experiment. The repeated study condition, on average, read through the text fourteen times, whereas the repeated recall condition only read the text three times. The benefit of repeated studying over retrieval practice on immediate tests results from the large discrepancy in the amount of exposure to the material between the two conditions. Critically, the amount of exposure to the material is not predictive of long-term learning, where the type of processing that students engage in during learning is a far more important.

Long-term retention is arguably more important for students in classes where knowledge is often needed days, weeks, months, or even years later. Therefore, learning activities must be compared on delayed assessments to better simulate authentic learning environments. Van Gog and Kester (2012) pitted the two learning activities against one another and examined problem-solving performance after a one-week delay. In their experiment, students learned how to troubleshoot electrical circuits by either studying four worked examples (i.e., EEEE) or by alternating between studying worked examples and problem solving (i.e., EPEP). Students who studied four worked examples performed numerically higher on the initial test and significantly better on the delayed test compared to students who had alternated between studying examples and solving practice problems. Unfortunately, the researchers included a confound in their experiment as they assessed performance at two time points for all participants – immediately after the learning phase and one-week later. Having the retention interval for the final test as a within-subjects factor makes it difficult to interpret delayed test performance because it is impossible to know whether any difference on the final test is due to the learning activity students engaged in or is due to differences in retrieval success on the first test. It is important to note that retrieval-based learning is helpful as long as students are able to successfully retrieve correct information (Karpicke et al., 2014) and studying four worked examples appears to have led to greater initial retrieval success during the learning phase, which provides an alternative explanation for the supposed benefit of the worked example condition on the final test.

In order to compare retrieval practice and worked examples without this confound, van Gog and colleagues (2015) repeated the experiment from van Gog and Kester (2012) but manipulated the retention interval as a between-subjects factor; students were either tested on their problem-solving ability 5 minutes after the learning phase or they completed the test one week later. Across four experiments, they found that alternating between studying worked examples and attempting to troubleshoot electrical circuit problems (i.e., the retrieval practice condition) produced equivalent long-term performance to the worked example condition. The researchers combined the results from these four experiments into a small-scale meta-analysis and concluded that retrieval-based learning was ineffective for highly complex, problem-oriented tasks based on a small, positive, but non-significant (d = 0.19) benefit that was in favor of solving practice problems over studying worked examples.

The lack of a retrieval practice effect for these problem-focused materials was taken as evidence that material complexity is a boundary for retrieval-based learning (van Gog & Sweller, 2015). In this literature, material complexity is defined in terms of element interactivity, meaning that materials are composed of individual elements that may exist in isolation or may interact with many other elements (Leahy et al., 2015; Sweller, 2010; Sweller et al., 1998). Less complex materials are said to be low in terms of element interactivity, meaning that each individual piece of information can be learned independently from other information that is present in the to-be-learned materials. An example of low element interactivity materials are word-pairs that are common in memory research as each pair does not require simultaneous processing of the other pairs in order to be learned. In contrast, complex materials are said to be high in terms of element interactivity, meaning that the materials are composed of a set of highly related ideas that must be processed simultaneously in working memory (van Gog et al., 2015). Problem-solving tasks that are commonly used in the worked example literature are often touted as the epitome of highly interactive materials because they contain multiple, related steps that build upon one another and must therefore be processed simultaneously in working memory (Chen et al., 2016; Leahy et al., 2015, Sweller, 2010; van Gog et al., 2015; van Gog & Sweller, 2015). These tasks impose a high cognitive burden on novice learners that inhibits learning, but this burden is alleviated when students study worked examples resulting in the worked example benefit.

Put simply, worked examples should lead to better problem-solving performance for complex tasks that are high in element interactivity. Yet, the small-scale meta-analysis conducted by van Gog and colleagues (2015) did not find the worked example effect. Leahy, Hanham, and Sweller (2015) also failed to find a benefit of studying worked examples over solving practice problems on a delayed test. In their study, children learned to read a bus schedule by either studying eight worked examples or by alternating between studying a worked example and solving a problem four times. In their Experiments 1 and 2, students who only studied worked examples performed better on an immediate test than students who completed example-problem pairs. However, the worked example effect disappeared when the final test occurred one week after learning. Hanham, Leahy, and Sweller (2017) also found a small, positive, but non-significant benefit of the problem-solving condition over the worked example condition when learning was assessed on a delayed test. The results from all of these experiments point toward the conclusion that retrieval-based learning activities are as effective, if not more effective, than studying worked examples for promoting long-term learning of problem-focused tasks.

A final comparison of the retrieval practice and worked example effects comes from Yeo and Fazio (2019). In their experiments, students learned how to solve Poisson probability problems by studying one worked example and solving three practice problems or by studying four worked examples. Yeo and Fazio (2019) found the worked example effect but only on immediate tests. There were no statistical differences between conditions on delayed tests for Experiments 1 and 3, meaning the worked example effect disappeared on delayed assessments. Experiment 2 differed from the other experiments in that students were learning with materials that had an identical problem format, differing only in the numerical values within the problem, and they were provided feedback in the form of the worked example after each practice problem. These changes increased initial problem-solving success relative to their other experiments, and the problem-solving condition outperformed the worked example condition on the delayed test. When considering their results along with the previous literature on this topic, Yeo and Fazio (2019) noted that the worked example effect tends to be found when problem-solving performance was low and disappeared when performance during learning was above 50%. This finding provides further evidence of the critical role that initial retrieval success plays in retrieval-based learning.

Small-scale, Exploratory Meta-analysis

A small-scale, exploratory meta-analysis was conducted to better understand the relationship between initial retrieval success and the effectiveness of retrieval practice relative to studying worked examples. Consequently, only studies that reported information about performance during the learning phase were included. A noteworthy observation about the worked example literature is that many studies do not report initial problem-solving performance, which appears to moderate the worked example effect (e.g., Chen et al., 2016; Darabi et al., 2007; Hanham et al., 2017; Leahy et al., 2015; van den Berge et al., 2013; van Gog et al., 2011). All future research should report performance during the learning phase for each practice problem. Additionally, only studies that assessed learning at a delay were included in the analysis because longer retention intervals better simulate authentic educational environments in which knowledge is often needed days, weeks, months, or even years later. Unfortunately, most of the research investigating the worked example effect has assessed learning on immediate tests, vastly limiting the number of included studies (Rawson, 2015; van Gog & Kester, 2012).

Table 1 presents the results from the small-scale, exploratory meta-analysis and has been intentionally sorted by initial performance to easily illustrate the strong, positive relationship between performance during the learning phase and the benefit of practicing retrieval over studying worked examples ($r = 0.85$, $p = .001$). Caution should be used when interpreting these results, and future research should complete a more exhaustive analysis of the worked example literature. This random effects meta-analysis model yielded an overall effect size of $d = 0.26$ [0.07, 0.46] that was in favor of the retrieval practice effect (see also Cumming, 2012; Smith &

Table 1

Effect sizes (g): retrieval practice vs. worked example

Initial Performance	Effect size [95% CI]	Publication	Experiment
28.30	-0.29 [-0.92, 0.33]	Yeo & Fazio (2019)	$E1_{iso}$
50.42	0.04 [-0.58, 0.66]	Van Gog et al. (2015)	$E1_{iso}$
53.87	0.04 [-0.45, 0.53]	Van Gog et al. (2015)	$E3_{id}$
64.89	-0.07 [-0.79, 0.64]	Van Gog et al. (2015)	$E2_{iso}$
64.89	0.45 [-0.25, 1.15]	Van Gog et al. (2015)	$E2_{id}$
69.12	0.34 [-0.19, 0.86]	Van Gog et al. (2015)	$E4_{iso}$
73.30	0.28 [-0.34, 0.90]	Yeo & Fazio (2019)	E3[a]$_{iso}$
75.00	0.19 [-0.43, 0.82]	Yeo & Fazio (2019)	$E3_{iso}$
77.00	0.28 [-0.22, 0.88]	Van Gog et al. (2015)	$E3_{iso}$
85.00	0.88 [0.23, 1.52]	Yeo & Fazio (2019)	$E2_{id}$
88.30	0.87 [0.22, 1.52]	Yeo & Fazio (2019)	E2[b]$_{id}$
Avg. initial performance 66.37%	**Overall effect size** g = 0.26 [0.07, 0.46]	--	--

Note. This table is intentionally sorted by initial performance during the learning phase to illustrate the strong positive relationship between initial retrieval success and the size of the retrieval practice effect relative to studying worked examples. Subscripts indicate the question type students were tested on in van Gog et al. (2015) or the type of problems in the learning phase in Yeo & Fazio (2019), iso = isomorphic and id = identical.

[a]The retrieval practice condition for this experiment instructed students to think back to the worked example before they were given the problem-solving test.

[b]The retrieval practice condition for this experiment involved feedback in the form of worked examples after each problem.

Karpicke, 2014; Smith, Roediger, & Karpicke, 2013; van Gog et al., 2015; Zamary & Rawson, 2018 for other examples of this type of analysis). Students who solved practice problems achieved higher scores on a delayed test compared to students who only studied worked examples. To further explore the relationship between initial performance and the benefit of retrieval practice, the studies were separated into two groups based on a median split of initial performance. There was a pronounced retrieval practice effect, $d = 0.47$ [0.19, 0.75], when examining the five studies that had initial performance above 69.12% ($N = 222$, $M = 79.72\%$). However, there was no benefit for retrieval practice, $d = 0.09$ [-0.15, 0.33], when looking at the six studies that had the lowest initial performance ($N = 263$, $M = 55.25\%$). Unexpectedly, there was not a benefit for learning by studying worked examples even when students struggled to solve practice problems. This analysis suggests that the worked example effect occurs only when worked examples are compared to conditions where students repeatedly fail to solve practice problems, and the failure to find the retrieval practice effect in previous experiments could be due to low performance during the learning phase in addition to poorly powered experiments.

Cognitive load theorists could argue that students who are performing well during learning are experts, and experts do not benefit from studying worked examples — a phenomenon referred to as the expertise reversal effect (Kalyuga, 2009; Kalyuga et al., 2003; Kalyuga et al., 2001). However, studies on expertise reversal often involve longitudinal designs spanning multiple sessions over several weeks with intensive training sessions (see Kalyuga, 2007, for the suggested experimental sequence). None of the experiments included in the small-scale, exploratory meta-analysis followed this design. Moreover, the average learning phase performance from the small-scale, exploratory analysis was 66.37%, which is hardly indicative of expertise. Nonetheless, proponents of cognitive load theory could argue that improving performance during the learning phase could afford expertise and therefore cognitive load theory would be correct in predicting an expertise reversal effect. If this path is taken, then the worked example effect is practically unimportant for educational purposes since it can be easily overturned when initial problem-solving success is supported.

Transfer Appropriate Processing: An Alternative Explanation

Theories of transfer-appropriate processing provide an alternative account for the relationship between initial problem-solving success and the effectiveness of retrieval-based learning. According to the idea of transfer-appropriate processing, optimal test performance depends on the compatibility between the cognitive processing that occurs during learning and the cognitive processing that will be required by the test. If learners are failing to solve practice problems during learning, then they are not engaging in the critical processing that the test will require. In this situation, worked examples provide a learning benefit because students are able to achieve a basic understanding of the necessary problem-solving procedure. The converse situation would occur when learners achieve sufficient problem-solving success during learning because they would be engaging in the processing that will be critical during the test, providing them with a unique advantage over the worked example condition.

Transfer-appropriate processing can also account for the absence of a retrieval practice effect when the retrieval-based learning activity required students to recall a previously studied worked example (van Gog et al., 2015; Yeo & Fazio, 2019). In these experiments, students were asked to recall everything they could remember about a previously studied worked example. Presumably, they retrieved key information that would help them solve future problems, such as the procedural steps or the needed mathematical operations. However, these instructions also led students to retrieve question specific information from the question prompt, such as the context of interest (e.g., the number of arriving and departing planes). When students are engaged in deep processing about the surface features of a practice problem, they are not necessarily engaging in the cognitive processing that will benefit them on a problem-solving test. Consequently, a retrieval practice effect for complex materials could be more likely if retrieval focused solely on the information that will be critical during the problem-solving test.

Overview of Experiments

Previous research has boldly claimed that retrieval-based learning benefits decrease (or even disappear altogether) as the complexity of the learning material increases (van Gog & Sweller, 2015). This claim is entirely premature because the supporting evidence is flawed. First, the studies that are put forth as evidence tend to find a small, positive benefit that is actually in favor of problem-solving over studying worked examples (van Gog et al., 2015). Second, experiments that have compared studying worked examples to practicing retrieval have been notably underpowered (Karpicke & Aue, 2015; Rawson, 2015). Third, the retrieval practice conditions often have abysmal initial learning performance without the provision of feedback (e.g., van Gog & Kester, 2012). Failing to solve problems is intuitively a poor learning strategy; a student would not be expected to improve his or her piano playing ability by listening to a melody once and failing to reproduce it repeatedly. Yet, this is exactly the comparison worked example studies make when attempting to evaluate the retrieval practice effect.

In order to address these major limitations in the current research base, three experiments were conducted to better inform educational practice. Experiment 1 investigated whether retrieval practice could improve memory for the procedural steps needed to solve complex statistical problems, Experiment 2 investigated whether enhanced memory for the procedural steps improved problem-solving performance, and Experiment 3 compared learning by studying worked examples, learning by solving practice problems without feedback, and learning by solving practice problems with feedback. In every experiment, students were asked to predict how well they would perform on the test (i.e., they made a judgment of learning) in order to investigate the influence of the learning activities on metacognition. Students were also asked to subjectively evaluate their learning activity based on their level of interest and engagement. The ideal learning activity would not only improve test performance but would also be engaging and interesting to students. Lastly, students rated how difficult the learning activity was and how much mental effort they had invested into the learning activity. This was critical information to collect because cognitive load theory has claimed that problem solving is difficult for novice learners and places a cognitive burden on them that inhibits their learning (van Gog & Kester, 2012; van Gog et al., 2015). These experiments not only hold theoretical significance as they tested competing predictions generated from the transfer-appropriate processing framework and cognitive load theory, but they also have important educational implications because they examined the potential limitations of retrieval-based learning for complex materials.

Experiment 1

The purpose of Experiment 1 was to demonstrate that retrieval practice can improve memory for the procedural steps that are needed to solve probability problems. In previous research, the retrieval practice condition required students to freely recall information about a previously studied worked example (van Gog et al., 2015; Yeo & Fazio, 2019). This type of processing has demonstrated an improvement in memory for the details included in the problem but does little to improve problem solving performance (Yeo & Fazio, 2019). Moreover, this retrieval practice activity appears to be a relatively weak form of an otherwise effective learning technique because students could be retrieving distracting and irrelevant information from the question prompt that would not benefit them on future problems. A more authentic application of retrieval practice would focus the mental processing that occurs during retrieval onto only the information that is relevant for solving problems. In Experiment 1, students were given a brief introduction about the Poisson probability distribution before studying or retrieving the procedural steps that are needed to solve these problems. After a brief delay, students were tested on their memory for the procedural steps.

Method

Subjects and design. Experiment 1 involved two between-subject conditions: a study only condition and a retrieval practice condition. The most recent meta-analysis of retrieval practice reported an effect size of g = 0.56 [0.51, 0.62] for same-day retention intervals (Adesope et al., 2017). Using this effect size, a power analysis indicated that 60 people were needed in each learning activity in order to achieve 85% power. One hundred and twenty Purdue University undergraduate students participated in exchange for course credit.

Materials. Students learned about the Poisson probability distribution by studying materials adapted from Yeo and Fazio (2019), which Lisa Fazio graciously shared with us. In their experiment, the Poisson probability problems were presented as word problems that were solved by applying a four-step solution; however, the fourth step of the procedure required students to identify the specific interval of interest and translate it into an inequality, apply and expand the equation for the Poisson probability distribution, and compute the probability. Due to differences in experimental design, the current experiment used a modified version of their materials. Namely, the last step for each problem was explicitly broken up into its separate components (see Figure 1). A general description of each step was also created to allow students to learn the procedure outside of the context for a specific problem (see Appendix B). These materials were ideal because they are unlikely to have been taught outside of college statistics courses, meaning students would most likely be novices learning how to solve complex, multi-step probability problems and should therefore experience cognitive load during learning.

Procedure. Students completed the 30-minute experiment online at a time and location of their choosing. After agreeing to participate in the experiment, students were asked if they had ever learned about the Poisson probability distribution and rated their memory for it on a scale from 0 to 10. All students were then informed that they would be learning how to solve probability problems for an upcoming test. They were first familiarized with the Poisson probability distribution by studying the formula sheet (see Appendix A), which explained the Poisson distribution and its formula, for 4 minutes and then by studying a worked example for 4 minutes (see Figure 1 for the worked example that all students studied). The worked example provided a detailed explanation of how to solve the problems by providing a general description

Worked Example 1

Suppose that the arrival and departure of airplanes at a domestic airport follow two independent Poisson distributions. In a one-hour period, it is expected on average that there are 4 arrivals and 3 departures. Find the probability that, in a randomly selected two-hour period, the airport handles 10 or more, but less than 13 arrivals and departures.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of arrivals in a two-hour period
1 hour → 4 arrivals
2 hours → $4 \times 2 = 8$ arrivals
So, $V_1 \sim P_o(8)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of departures in a two-hour period
1 hour → 3 departures
2 hours → $3 \times 2 = 6$ departures
So, $V_2 \sim P_o(6)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of arrivals and departures in a two-hour period
$T \sim P_o(8 + 6)$
$T \sim P_o(14)$

Step 4: Determine the boundaries of X

10 or more and less than 13
$(10 \leq X < 13)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(10 \leq X < 13) = P(X = 10) + P(X = 11) + P(X = 12)$

$$= \frac{e^{-14} \cdot 14^{10}}{10!} + \frac{e^{-14} \cdot 14^{11}}{11!} + \frac{e^{-14} \cdot 14^{12}}{12!}$$

Step 6: Calculate the probability

= .249

Figure 1. The worked example that all conditions studied during the beginning of the learning phase in every experiment.

for each step, the meaning for each step in the current problem context, and the corresponding mathematical operations that were computed for each step.

The remainder of the learning phase differed depending on the learning activity that students were randomly assigned to. Those assigned to the study only condition studied the procedural steps for a total of 10 minutes and 30 seconds (see Appendix B). Those assigned to the retrieval practice condition alternated between studying the procedural steps for 1 minute and 30 seconds and recalling them from memory for 2 minutes for a total of 10 minutes and 30 seconds (i.e., SRSRSR). After finishing the learning phase of the experiment, students predicted how well they would be able to remember the procedural steps on a test in a few minutes and made their ratings on a scale from 0% to 100% in increments of 10 (0, 10, … 90, 100) by clicking a radio button displayed on the screen. They also rated how engaging, difficult, and interesting their learning activity was (adapted from Blunt & Karpicke, 2014) on a scale from 0 to 10 in increments of 1 (0, 1, ... 9, 10) as well as the amount of mental effort (Paas, 1992) they invested in the learning activity on a scale of 1 to 9 in increments of 1 (1, 2, … 8, 9). Students then played Pac-Man as a distractor task for 4 minutes before they completed a final 4-minute recall of the procedural steps.

Results

The data and the analysis script for each experiment is available at garrettoday.info/projects

Scoring. Recall responses of the procedural steps were scored by two independent raters. Each step in the procedure was scored, and raters awarded 1 point for fully correct responses, 0.5 points for partially correct responses, or 0 points for incorrect responses. The raters agreed on 90% of the responses, and the scores were averaged across raters for the purpose of analyses.

Recall performance during learning. Table 2 shows recall performance during the learning phase for students who practiced retrieval. An initial analysis indicated that some of the procedural steps were easier to remember than others, but step performance did not interact with recall number. Since memory for all of the procedural steps would be needed to successfully solve problems, results have been collapsed across the procedural steps. Consistent with the vast literature on retrieval practice and our predictions, students' memory for the procedural steps improved across recall attempts with interspersed restudy opportunities, $F(2, 118) = 83.26$, $p < .001$, $\eta_p^2 = .59$.

Final recall performance. Figure 2 compares final recall performance between students that practiced retrieval during learning and those that only studied the procedural steps. An initial analysis indicated that some of the procedural steps were easier to remember than others, but step performance did not interact with the learning activity. Since memory for all of the procedural steps would be needed to successfully solve problems, results have been collapsed across the procedural steps. Students that practiced retrieval to learn the procedural steps for solving Poisson probability problems outperformed students who simply studied the procedural steps with an overall benefit of 25%, $d = 0.86$, [0.49, 1.23].

Subjective ratings. There was no difference in the proportion of students who had prior knowledge about the Poisson probability distribution in the retrieval practice condition (.20) and the study only condition (.17, $p = .81$, Fisher's exact test). Additionally, students' self-reported memory for the Poisson probability distribution was very low ($M = 0.58$ out of 10) and did not differ between the conditions, $d = 0.16$ [-0.20, 0.52]. Table 3 shows students' subjective ratings of mental effort, difficulty, interest, and engagement as well as their metacognitive predictions.

Table 2

Proportion of the procedural steps correctly recalled during retrieval practice

	Recall 1	Recall 2	Recall 3
Experiment 1	.27 (.24)	.61 (.26)	.66 (.28)
Experiment 2	.43 (.22)	.66 (.21)	.77 (.22)

Note. Standard deviations are reported in parentheses.

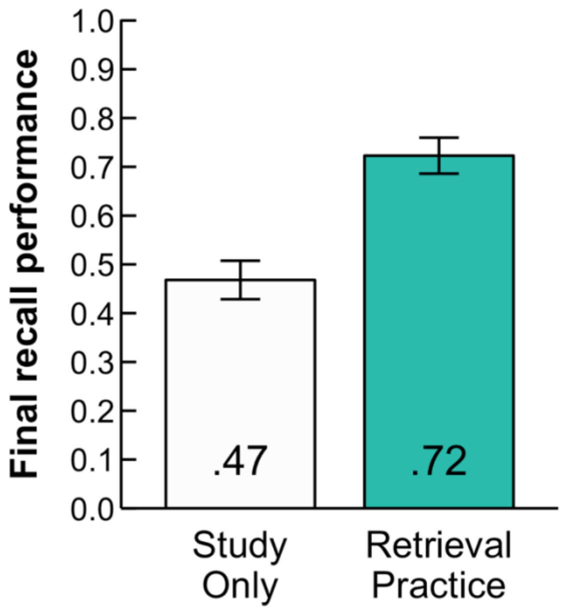

Figure 2. Proportion of the procedural steps correctly recalled on the test in Experiment 1. Error bars represent standard errors of the means.

Table 3

Subjective ratings for the learning activities in Experiments 1, 2, and 3

	Mental Effort	Difficulty	Interest	Engagement	JOL
Experiment 1					
Study Only	5.28 (1.94)	5.33 (2.43)	3.13 (2.67)	3.63 (2.62)	58.17 (28.49)
Retrieval Practice	5.48 (2.10)	5.60 (2.48)	2.97 (2.77)	4.82 (3.02)	58.00 (26.54)
Experiment 2					
Study Only	4.93 (1.55)	4.23 (2.37)	2.92 (2.47)	3.97 (2.49)	66.53 (25.23)
Retrieval Practice	5.51 (2.06)	4.41 (2.36)	4.45 (2.37)	6.05 (2.41)	71.33 (21.89)
Experiment 3					
Worked Examples	4.06 (1.92)	2.39 (1.73)	4.38 (2.72)	5.49 (2.72)	86.64 (13.92)
Problem Solving	4.60 (2.07)	3.75 (2.38)	5.09 (2.75)	6.78 (2.71)	75.06 (24.14)
Problem Solving w/ Feedback	4.96 (2.31)	3.76 (2.23)	5.53 (2.33)	6.87 (2.23)	82.38 (15.99)

Note. Difficulty, interest, and engagement were reported on a scale from 0 to 10. Mental effort was reported on a scale from 1 to 9, consistent with prior literature (Paas, 1992). Judgments of learning were reported on a scale from 0 to 100. Standard errors of the means are shown in parentheses.

Students in both conditions made similar judgments of learning, *d* = -0.01 [-0.36, 0.35]. However, students in the study only condition displayed significant overconfidence, predicting that they would remember 11% more of the material than they ultimately would on the final test. Students who had practiced retrieval were underconfident in their metacognitive predictions by about 14%. The subjective ratings for the learning activities did not differ between the conditions in terms of difficulty, *d* = 0.11 [-0.25, 0.47], interest, *d* = -0.06 [-0.42, 0.30], or mental effort, *d* = 0.10 [-0.26, 0.46]. Interestingly, students who used retrieval practice reported more engagement than students who only studied the procedural steps, *d* = 0.42 [0.06, 0.78].

Discussion

The purpose of Experiment 1 was to demonstrate that retrieval practice could be specifically targeted to improve memory for the procedural steps that are needed to solve complex probability problems. Compared to the study only condition, retrieval practice greatly improved retention of the procedure when learners were tested a few minutes later. Learners who practiced retrieval also reported higher levels of engagement during learning and were less overconfident in their metacognitive predictions than those who only studied the procedural steps. Notably, these learning benefits did not come at a cost of increased mental effort or cognitive load.

Experiment 2

The results from Experiment 1 indicated that retrieval practice is effective at improving memory for the procedural steps that need to be followed when solving Poisson probability problems. At first glance, these results appear to provide evidence against cognitive load theory and its concept of element interactivity, which asserts that retrieval practice is ineffective with highly complex and related materials. However, the results from Experiment 1 could be interpreted as being compatible with this view if the procedural steps are deemed to be low in terms of complexity or element interactivity (i.e., each step could be learned in isolation from the other steps). Therefore, Experiment 2 extended the results from Experiment 1 by using a final test that was intended to impose a greater demand on working memory — a problem solving test. Certainly, a problem-solving test that requires students to identify key information contained in the question prompt, remember and apply the procedural steps that build toward the solution, and perform the necessary algebraic calculations in order to compute the probability is a more complex and interactive assessment than verbatim memorization of the procedural steps. In fact, the primary goal of the worked example literature, which gave rise to the idea of element interactivity, has been to increase problem solving performance. According, to cognitive load theory and its concept of element interactivity, no benefit of retrieval practice is predicted for the problem-solving test. Alternatively, if successful problem solving requires sufficient memory for the procedural steps then retrieval practice should improve memory for those steps and should therefore facilitate problem-solving performance compared to the study only condition. This pattern of results would provide additional evidence that the retrieval practice effect is alive and well with complex materials.

Method

Subjects and design. Experiment 2 involved two between-subject conditions, a study only condition and a retrieval practice condition that were both tested on their ability to solve problems. The lower bound (d = 0.49) of the effect size from Experiment 1 was chosen to serve as a conservative effect size for determining the needed sample size. Using this effect size, a power analysis indicated that 75 people were needed in each learning activity in order to achieve 85% power. One hundred and fifty Purdue University undergraduate students participated in exchange for course credit.

Materials. Experiment 2 used the same materials as Experiment 1, which were adapted and modified by Yeo and Fazio (2019). The primary change from Experiment 1 is that the final assessment was an eight-item problem-solving test (see Appendix E). Two of the problems on the test were isomorphic (i.e., required the same problem-solving procedure but had a different question prompt, setting, and values) to the worked example that students studied at the beginning of the learning phase. The remaining six questions were designed to test students' ability to transfer their understanding to problems that required a subset or variation of the procedure they had learned. The problems were presented in the same order for all students, following the procedure from Yeo and Fazio (2019). Since the final assessment in this experiment involved probability calculations, students were provided with a TI-30XS calculator. Consequently, instructions on how to use the calculator were added to the bottom of the formula sheet, which explained the Poisson distribution, provided the formula for it, and was present during the test (see Appendix A).

Procedure. Experiment 2 consisted of a single, laboratory-based experimental session that lasted 64 minutes. Students were tested in small groups up to five people. The learning phase of the experiment was completed on a computer, and the test phase of the experiment was completed with pen and paper. After agreeing to participate in the experiment, students were asked if they had ever learned about the Poisson probability distribution before and rated their memory for it on a scale from 0 to 10. Students were then informed that they would be learning how to solve probability problems for an upcoming test. They were first familiarized with the Poisson probability distribution by studying the formula sheet (see Appendix A) for 4 minutes and then by studying a worked example for 4 minutes. The worked example provided a detailed explanation of how to solve the problems by providing a general description for each step, the meaning for each step in the current problem context, and the corresponding mathematical operations that were computed for each step (see Figure 1).

The remainder of the learning phase differed depending on the learning activity that students were randomly assigned to. Those assigned to the study only condition continued studying the procedural steps for a total of 10 minutes and 30 seconds. Those assigned to the retrieval practice condition alternated between studying the procedural steps for 1 minute and 30 seconds and recalling them from memory for 2 minutes for a total of 10 minutes and 30 seconds (i.e., SRSRSR). After finishing the learning phase of the experiment, students predicted how well they would do on a problem-solving test in a few minutes, and they made their ratings on a scale from 0% to 100% in increments of 10 (0, 10, … 90, 100) by clicking a radio button displayed on the screen. They also rated how engaging, difficult, and interesting their learning activity was (adapted from Blunt & Karpicke, 2014) on a scale from 0 to 10 in increments of 1 (0, 1, ... 9, 10) as well as the amount of mental effort (Paas, 1992) they invested in the learning activity on a scale of 1 to 9 in increments of 1 (1, 2, … 8, 9). Students then played Pac-Man as a distractor task for 4 minutes.

Students were then handed a pen, a calculator, a formula sheet, and a test packet. They were informed that they would be given 5 minutes to solve each problem. One problem was presented on each page of the test packet. If they finished a problem early, they were told to double-check their work for that problem only. When the time limit for the current problem was reached, the experimenter instructed all students to turn to the next problem. The experimenter carefully monitored time and compliance with the instructions.

Results

The data and the analysis script for each experiment is available at garrettoday.info/projects

Scoring. Two independent raters scored the recall responses in the same manner as Experiment 1. Each step in the procedure was scored, and raters awarded 1 point for fully correct responses, 0.5 points for partially correct responses, or 0 points for incorrect responses. The raters agreed on 95% of the responses, and the scores were averaged across raters for the purpose of analyses. The problem-solving test was scored by awarding 1 point for each correct solution. Responses that included an apparent computation error (e.g., 10 + 33 = 45) but provided an otherwise correct response were awarded full credit. Twenty percent of the responses were scored by two independent raters. The raters agreed on 96% of the responses, discrepancies between the raters were resolved through discussion, and the remaining responses were scored by one rater.

Recall performance during learning. Table 2 shows recall performance during the learning phase for students who practiced retrieval. As with Experiment 1, an initial analysis indicated that some of the procedural steps were easier to remember than others, but step performance did not interact with recall number. Consequently, the results have been collapsed across steps to better represent learners' overall understanding of the procedure. Consistent with the vast literature on retrieval practice and the results from Experiment 1, students' memory for the procedural steps improved across retrieval attempts with interspersed restudy opportunities, $F(2, 148) = 94.52$, $p < .001$, $\eta_p^2 = .56$.

Problem solving test performance. Figure 3 shows performance on the problem-solving test that occurred 4 minutes after the learning phase and was composed of two near transfer and six far transfer questions. There were no statistically significant differences between the learning activities on near ($d = 0.00$ [-0.32, 0.32]) or far transfer questions ($d = -0.19$ [-0.51, 0.13]). Retrieval practice of the procedural steps did not produce a learning benefit on a problem-solving test compared to the study only condition. Additionally, the relationship between recall and problem-solving performance was rather small ($r = .26$, $p = .03$). Taken together, improved memory for the procedural steps via retrieval practice did not facilitate problem solving performance, suggesting that other types of processing are critical for solving these types of problems.

Subjective ratings. There was no difference in the proportion of students who had prior knowledge about the Poisson probability distribution in the retrieval practice condition (.16) and the study only condition (.15, $p = 1.00$, Fisher's exact test). Additionally, students' self-reported memory for the Poisson probability distribution was very low ($M = 0.53$ out of 10) and did not differ between the conditions, $d = 0.01$ [-0.31, 0.33]. Table 3 shows students' subjective ratings of mental effort, difficulty, interest, and engagement as well as their metacognitive predictions. Students in both conditions made similar judgments of learning, $d = 0.20$ [-0.12, 0.52]. Their subjective ratings did not differ between the learning activity conditions in terms of difficulty ($d = 0.08$ [-0.24, 0.40]) or mental effort, ($d = 0.29$ [-0.04, 0.61]). However, students who engaged in

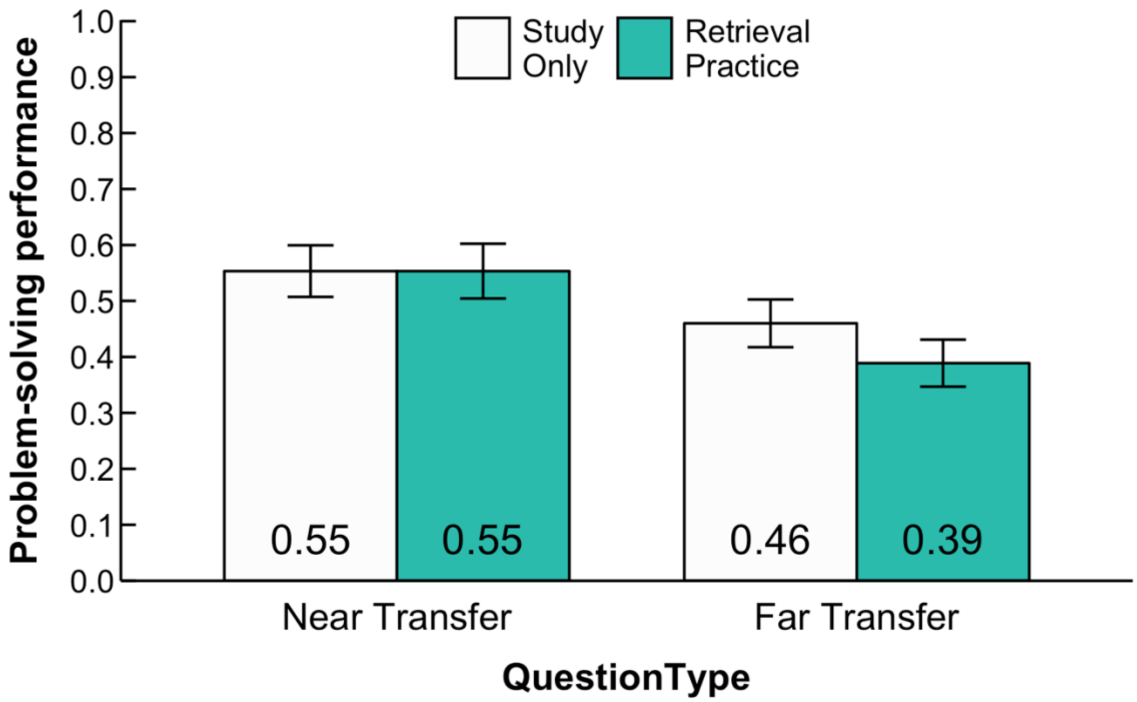

Figure 3. Problem-solving test performance for near and far transfer questions in Experiment 2. Error bars represent standard error of the means

retrieval practice reported higher levels of engagement (*d* = 0.85 [0.51, 1.18]) and interest (*d* = 0.63 [0.31, 0.96]) during learning than students who only studied the procedural steps.

Discussion

The results from Experiments 1 and 2 appear to support cognitive load theory and its concept of element interactivity in that a retrieval practice benefit was observed in Experiment 1 when the learning assessment was to recall the procedural steps and could be labeled as low in terms of material complexity. The retrieval practice effect then vanished in Experiment 2 when the learning assessment imposed a greater working memory demand; students had to identify key information contained in the question prompt, remember and apply the procedural steps that build toward the solution, and perform the necessary algebraic calculations in order to solve the problems. Put simply, retrieval practice produced a large benefit when the assessment imposed low working memory demands, but this benefit disappeared when students were given a problem-solving test that induced a greater burden on working memory. As predicted by cognitive load theory and its concept of element interactivity, material complexity appears to be a boundary condition for the retrieval practice effect when retrieval is involved as repeated recall of the procedural steps. Nevertheless, an alternative explanation could account for the same pattern of data and would generate testable predictions that would be inconsistent with the predictions from cognitive load theory. This alternative explanation draws on transfer-appropriate processing, the idea that the relationship between the type of processing students engage in during learning and the type of processing required when tested determines the memorability of information (Karpicke; 2017, Morris, Bransford, & Franks, 1977; Roediger, 1990).

According to transfer-appropriate processing, for learners to achieve optimal test performance they need to study in such a way that they are engaging in the same mental processing that will be required of them when tested. The retrieval practice condition in Experiment 1 accomplished this by having the students engage in free recall, the exact mental task that would be required of them during the test. As a result, retrieval practice uniquely afforded test compatible processing during learning and produced superior final recall performance relative to the study only condition, which did not afford the opportunity for students to engage in test compatible processing. However, Experiment 2 used a problem-solving test that required students to identify key information contained in the question prompt, remember and apply the procedural steps that build toward the solution, and perform the necessary algebraic calculations in order to solve the problems. In this experiment, the retrieval practice condition did not produce better test performance compared to the study only condition, which could be due to the retrieval practice condition no longer affording the critical test compatible processing.

Experiment 3

Experiment 3 tested the competing predictions from cognitive load theory and the transfer-appropriate processing framework. According to cognitive load theory and its concept of element interactivity, retrieval practice is ineffective for complex materials and studying worked examples is the optimal study strategy for novice learners. Alternatively, the transfer-appropriate processing account predicts that solving practice problems would afford the same cognitive processing that would be critical on a problem-solving test, and this learning-test compatibility would facilitate test performance. The caveat with this prediction is that students must be engaging in the test compatible cognitive processing during learning, which would be demonstrated by students solving some of the practice problems correctly. If learners are unable to solve practice problems successfully then they are most likely not be engaging in test compatible processing and would not have an advantage on the test.

In this experiment, one group of students solved practice problems, one group of students solved practice problems and received feedback, and one group of students only studied worked examples. Feedback was provided to one group of students in order to enhance performance during the learning phase in hopes of ensuring that this group of students would engage in the cognitive processing that would be required on the problem-solving test. Transfer-appropriate processing would then predict that solving practice problems with feedback would result in the highest levels of test performance because it would afford the most test compatible processing during learning. Conversely, cognitive load theory predicts that studying worked examples would produce the best test performance because the retrieval practice effect is ineffective for complex materials that are high in element interactivity.

Method

Subjects and design. Experiment 3 used three between-subject conditions: worked examples only, problem solving only, and problem solving with feedback. The primary comparison of interest is between the worked example condition and the problem solving with feedback condition because feedback was predicted to improve initial problem-solving success, increasing the probability that students would be able to engage in test compatible cognitive processing during learning. Two meta-analyses have estimated the effect size of retrieval practice with feedback to be $g = 0.63$, 95% CI [0.58, 0.68] and $g = 0.73$, 95% CI [0.61, 0.86] (Adesope et al., 2017; Rowland, 2014, respectively). However, a far more conservative effect size of $d = 0.47$ was chosen for determining the needed sample size because past research has argued that the retrieval practice effect is smaller for complex, problem-oriented materials (e.g., van Gog & Sweller, 2015, but see Karpicke & Aue, 2015 for an alternative view). Note that this value is well below the lower bound of both meta-analyses and is the effect size from small-scale meta-analysis that focused on studies that had initial performance above 69.12%. Using this effect size, a power analysis indicated that 100 people were needed in each learning activity in order to achieve 90% power. Three hundred Purdue University undergraduate students participated in exchange for course credit. This large sample will provide far more power than has been present in previous research that has compared retrieval practice and worked examples.

The experiment involved two experimental sessions, spaced one week apart. In Session 1, students learned how to solve statistical problems about the Poisson distribution. Students returned one week later for Session 2, which involved a final problem-solving test. Session 1 was completed in 30 minutes and Session 2 was completed in 55 minutes.

Materials. Experiment 3 used the same materials as the previous experiments, which were adapted and modified from Yeo and Fazio (2019), but the learning phase focused on problem solving rather than focusing on memorization of the general procedure.

***Practice problems*.** For the learning phase of the experiment, students encountered four Poisson probability problems, each requiring a six-step solution. All students, regardless of condition, initially studied a worked example (see Figure 1). Depending on the learning activity that students were randomly assigned to, the three other problems were either studied as a worked example (see Appendix C) or were presented as practice problems that needed to be solved (see Appendix D). These three problems were isomorphic to the initial worked example, meaning they had the same basic problem structure but different surface features (i.e., a different question prompt, setting, and values).

***Test problems*.** For the test phase of the experiment, students solved two near transfer problems that were isomorphic to those encountered during the learning phase, six far transfer problems, and three questions that were repeated from the learning phase (i.e., Practice Problems 2, 3, and 4). The near transfer problems closely followed the format of the problems encountered during the learning phase. The far transfer problems differed from the near transfer problems by requiring a subset or variation of the problem-solving procedure. The problems were presented in the same order for all students, following the procedure from Yeo and Fazio (2019) with the modification that the last three questions from the learning phase were included after the eighth test question (see Appendix E).

Procedure. The learning phase of the experiment lasted 30 minutes, and students were tested in small groups up to five people. After agreeing to participate in the experiment, students were asked if they had ever learned about the Poisson probability distribution before and rated their memory for it on a scale from 0 to 10. Everyone was familiarized with the Poisson probability distribution by studying the formula sheet for 4 minutes (see Appendix A), which explained the Poisson distribution, its formula, and provided instructions on how to use the calculator. Students had access to the formula sheet for the entirety of the learning phase to ensure that they were concentrating on learning the procedure itself rather than simply memorizing the formula. Importantly, the formula sheet did not contain information on how to solve the upcoming problems.

All students, regardless of condition, then studied the same worked example for 6 minutes (see Figure 1). Worked examples provided a detailed explanation of how to solve the problems by providing a general description of each step, the meaning of each step in the current problem context, and the corresponding mathematical operations that were computed for each step. The remainder of the learning phase differed depending on the learning activity that the student was randomly assigned to. Students in the worked example condition studied three isomorphic worked examples for 6 minutes per problem (S_2 S_3 S_4). Students in the problem solving only condition attempted to solve the three isomorphic or near transfer problems and were given 6 minutes per problem (P_2 P_3 P_4). Rather than alternating between solving a practice problem and then studying an entirely new worked example (e.g., van Gog et al., 2015), students in the problem solving with feedback condition were provided with the worked example for each of the practice problems they attempted. This procedure more accurately resembles the restudy phase in traditional retrieval practice experiments, where students are given an opportunity to review the exact information they had attempted to remember (Roediger & Karpicke, 2006). Studying a completely new example requires students to spend time reading and understanding the new demands of a question. Feedback for the exact practice problem they had just attempted was hypothesized to be more effective because it would allow students to find their mistakes quicker as they would already be familiar with the problem, granting them more time to hone their problem-solving knowledge. Put simply, students in the problem solving with feedback condition attempted to solve three isomorphic problems but received feedback after 4 minutes in the form of the worked example for that particular problem, which they studied for 2 minutes (P_2E_2 P_3E_3 P_4E_4). All students saw the same problems, for the same amount of time, in the same order. Time was carefully tracked by the researcher, who collected and then handed out each practice problem or worked example.

After finishing the learning phase of the experiment, students predicted how well they would do on a problem-solving test one week later, and they made their ratings on a scale from 0% to 100% in increments of 10 (0, 10, … 90, 100). They also rated how engaging, difficult, and interesting their learning activity was (adapted from Blunt & Karpicke, 2014) on a scale from 0 to 10 in increments of 1 (0, 1, ... 9, 10) as well as the amount of mental effort (Paas, 1992) they invested in the learning activity on a scale of 1 to 9 in increments of 1 (1, 2, … 8, 9).

When students arrived one-week later for problem-solving test, they were then handed a pen, a calculator, a formula sheet, and a test packet. One problem was presented on each page of the test packet. If they finished a problem early, they were told to double-check their work for that problem only. When the time limit for the current problem was reached, the experimenter instructed all students to turn to the next problem. The experimenter carefully monitored time and compliance with the instructions. The test consisted of two near transfer, six far transfer, and three repeated questions for a total of 11 questions. Session 2 lasted 55 minutes.

Results

The data and the analysis script for each experiment is available at garrettoday.info/projects

Scoring. All responses to the problems either during learning or at test were scored by awarding 1 point for each correct solution. Responses that included an apparent computation error (e.g., 10 + 33 = 45) but provided an otherwise correct response were awarded full credit. Twenty percent of the responses were scored by two independent raters. The raters agreed on 92% of the responses, discrepancies between the raters were resolved through discussion, and the remaining responses were scored by one rater.

Problem-solving performance during learning. Table 4 reports performance on the practice problems that students in the retrieval practice conditions attempted to solve during the learning phase. Contrary to our predictions, the problem-solving condition had higher performance on the first (d = 0.44 [0.16, 0.72] and second (d = 0.42 [0.14, 0.70]) practice questions compared to the problem solving with feedback condition. However, at the end of the learning phase the two problem solving conditions performed equally, d = 0.00 [0-.28, .28]. The most reasonable explanation for the initial difference between the groups is that the problem-solving only condition had 6 minutes to solve each problem and the problem solving with feedback condition had only 4 minutes to solve each problem, which was done in order to maintain equivalent total time per problem between the conditions. Looking across performance on the practice problems, solving problems with feedback produced a positive learning trajectory (an increase of .27 from Problem 2 to 4), whereas the problem-solving condition led to little growth over the learning phase (an increase of .07 from Problem 2 to 4). It is possible that if the learning phase had included more learning opportunities then the problem solving with feedback condition would have continued to improve. Future research should explore this possibility.

Table 4

Problem-solving performance during learning in Experiment 3

	Problem 2	Problem 3	Problem 4
Problem Solving	.73 (.45)	.78 (.42)	.79 (.41)
Problem Solving w/ Feedback	.52 (.50)	.59 (.49)	.79 (.41)

Note. Standard deviations are reported in parentheses.

Problem-solving test performance. Figure 4 displays the performance for each type of problem on the problem-solving test that occurred one week after the learning phase. Contrary to predictions from cognitive load theory, the worked example benefit was absent for all types of test problems. Solving practice problems without feedback led to a small, non-significant benefit over studying worked examples for repeated (*d* = 0.16 [-0.12, 0.44]), near transfer (*d* = 0.13 [-0.15, 0.41]), and far transfer questions (*d* = 0.06 [-0.21, 0.34]). These results are consistent with previous studies that have found limited benefits of solving practice problems over studying worked examples on delayed assessments (Hanham et al., 2017; Leahy et al., 2015; van Gog et al., 2015; Yeo & Fazio, 2019). Solving these complex probability problems with little to no guidance produced equivalent problem-solving test performance compared to studying worked examples. More importantly, is the critical comparison between the problem-solving with feedback and the worked example conditions. Students that solved practice problems and received feedback outperformed the worked example condition on the test problems that were repeated from the learning phase, *d* = 0.40 [0.12, 0.68], and on the near transfer problems, *d* = 0.34 [0.06, 0.62], but not on far transfer problems, *d* = 0.03 [-0.25, 0.31]. Put simply, problem solving without guidance led to slightly better test performance than studying worked examples, and this problem-solving benefit was enhanced when feedback was provided.

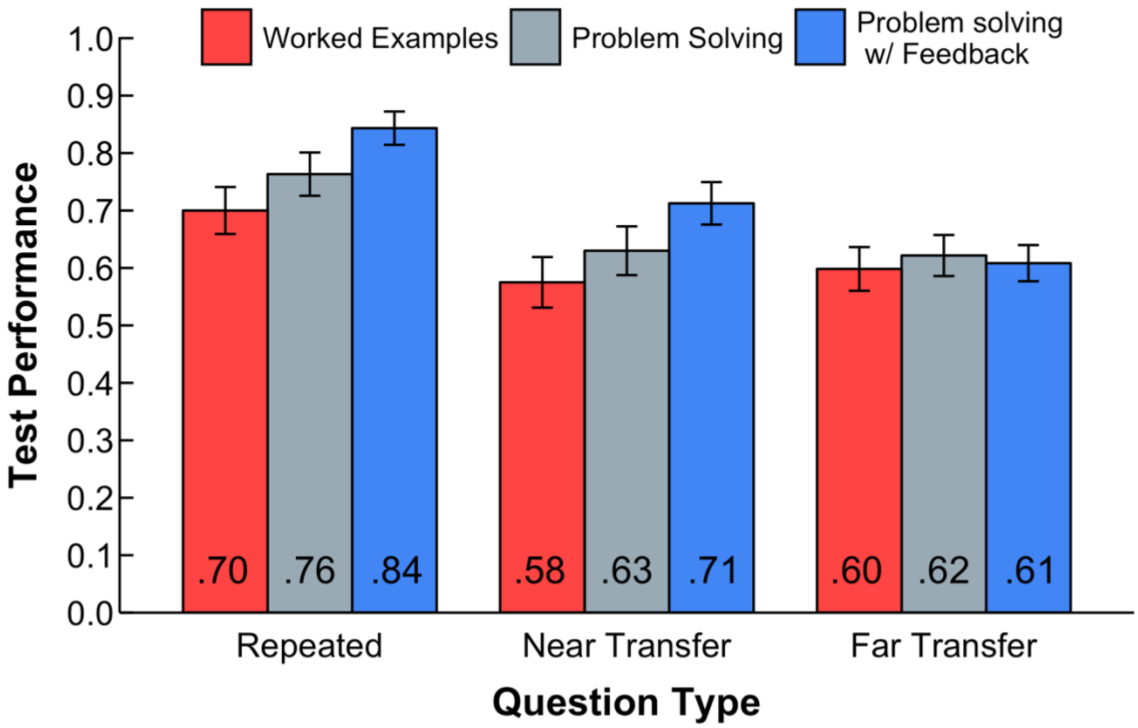

Figure 4. Problem-solving test performance for repeated, near transfer, and far transfer questions in Experiment 3. Error bars represent standard errors of the means.

Exploring the relationship between practice problem and test problem performance. Since initial learning performance moderates the benefit from retrieval-based learning (Rowland, 2014), an exploratory analysis was conducted to better examine the importance of practice problem performance. An initial correlation was computed between practice problem performance and overall test performance, revealing a strong, positive relationship (r = .50, p < .001). Further analysis examined test performance based on the number of practice questions students solved correctly during the learning phase. Table 5 provides the proportion of students who solved none, one, two, or three of the practice questions correctly for each of the problem-solving conditions. Interestingly, 63% of students in the problem-solving condition answered every practice question correctly compared to only 32% in the problem solving with feedback condition. As previously stated, the problem-solving condition had more time devoted to solving problems and this additional time appears to have led to higher problem-solving performance. Of particular interest is that many of the students in these conditions experienced repeated failure during the learning phase solving less than 33% of the practice problems correctly.

Figure 5 displays the average overall test performance for each problem-solving condition separated by the number of practice problems solved successfully, and the dashed line represents average test performance for the worked example condition, M = .62. These results should be interpreted with caution because the sample has been divided based on initial performance, which greatly reduces power, and did not involve random assignment. Nevertheless, the relationship between initial performance and the worked example effect is striking. When students are solving problems without feedback and fail to solve any of the practice problems correctly, they have extraordinarily low-test performance leading to a sizeable benefit for worked examples of 53%, d = 2.65 [1.95, 3.34]. This pattern is consistent with the

Table 5

Proportion of students that solved 0, 1, 2, or 3 of the practice problems correctly in Experiment 3

	0 correct	1 correct	2 correct	3 correct
Problem Solving	.12	.09	.16	.63
Problem Solving w/ Feedback	.13	.16	.39	.32

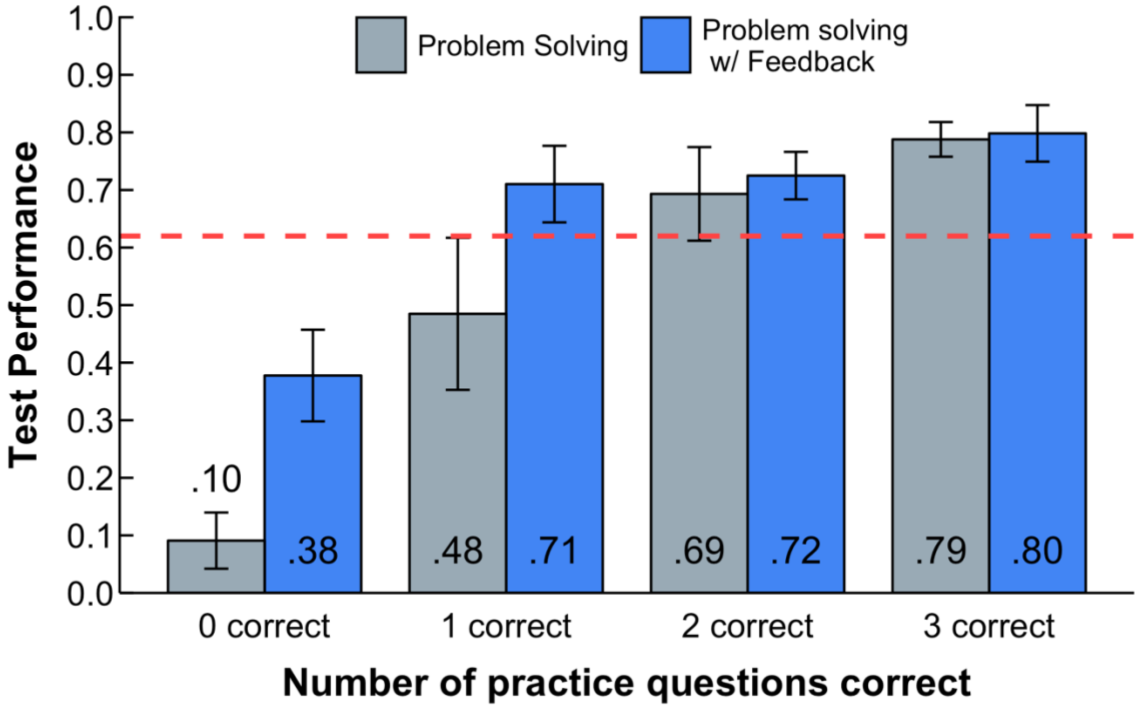

Figure 5. Problem-solving test performance in Experiment 3 based on initial problem-solving performance. The dashed read line represents the overall test performance of the worked example condition (.62). Error bars represent standard errors of the means.

idea of transfer-appropriate processing as these students are not engaging in the mental processing during learning that will benefit them when tested and subsequently perform poorly on the problem-solving test. The worked example effect was drastically reduced (14%, d = 0.35 [-0.34, 1.03]) when students were able to solve one of the practice problems correctly in the absence of feedback. There was a slight benefit for solving problems without feedback when students solved two practice problems correctly (7%, d = 0.22 [-0.31, 0.74]) and an even bigger benefit in favor of problem solving when students solved all three practice problems correctly (17%, d = 0.51 [0.19, 0.83]).

Feedback was particularly beneficial for students who struggled on the practice problems. Compared to the problem solving only condition, failing to solve every practice problem and receiving feedback cut the worked example benefit in half (24%, d = 0.82, [0.23, 1.41]). When students solved one (9%, d = 0.31 [-0.22, 0.84]) or two (10%, d = 0.35 [-0.02, 0.72]) of the practice problems correctly and received feedback there was a small retrieval practice effect. Finally, there was a large benefit of retrieval-based learning over studying worked examples when students were able to solve every practice problem correctly and received feedback (18%, d = 0.58 [0.18, 0.99]). Although not definitive, this exploratory analysis provides further evidence that initial retrieval success is a critical component of retrieval-based learning.

Subjective ratings. There were no differences in the number of students who had prior knowledge about the Poisson probability distribution between the conditions (p = .67, Fisher's exact test). Additionally, students' self-reported memory for the Poisson probability distribution was very low (M = 0.57 out of 10) and did not differ between the conditions, $F(2, 297) = 0.41$, p = .66, $\eta_p^2 = 0.03$. Table 3 shows students' subjective ratings of mental effort, difficulty, interest, and engagement as well as their metacognitive predictions. Students in the problem solving with feedback condition and problem-solving condition provided similar ratings for difficulty (d = 0.01 [-0.27, 0.28]), interest (d = 0.17 [-0.11, 0.45]), engagement (d = 0.03 [-0.24, 0.31]), and mental effort (d = 0.16 [-0.12, 0.44]). Solving problems with feedback led to higher judgments of learning than solving problems without feedback (d = 0.36 [0.08, 0.64]). Worked examples led to higher judgments of learning than the problem solving with feedback (d = 0.28 [0.01, 0.56]) and the problem-solving condition (d = 0.59 [0.30, 0.87]). The problem-solving (d = 0.65 [0.37, 0.93]) and problem solving with feedback (d = 0.69 [0.40, 0.97]) conditions were rated as more difficult than the worked example condition. However, these more difficult conditions were also rated as more engaging (problem solving: d = 0.48 [0.20, 0.76]; problem solving with feedback: d = 0.56 [0.27, 0.84]) and more interesting (problem solving: d = 0.26 [-0.02, 0.54]; problem solving with feedback: d = 0.45 [0.17, 0.74]) than studying worked examples. Importantly, the worked example condition reported lower ratings of mental effort than either the problem-solving (d = 0.27 [-0.01, 0.55]) or problem solving with feedback conditions (d = 0.42 [0.14, 0.70]).

Discussion

The purpose of Experiment 3 was to test the competing predictions from cognitive load theory against predictions from the transfer-appropriate processing framework by having students learn to solve Poisson probability problems by either studying worked examples or by solving practice problems. The results from the problem-solving test conflict with the predictions from cognitive load theory because the worked example effect was not observed. In fact, the worked example condition produced the lowest performance for all types of test problems. The results were instead consistent with the predictions that were generated from the transfer-appropriate processing framework because students who solved practice problems with feedback achieved the highest scores on the delayed test, providing additional evidence that the retrieval practice effect is alive and well with complex materials.

Experiment 3 also demonstrated the usefulness of worked examples as a method for delivering feedback. Students that studied worked examples as feedback after attempting to solve practice problems performed the best on repeated and near transfer questions. These students also reported the highest levels of engagement and interest during learning. Moreover, the exploratory analysis illustrated in Figure 5 suggested that this form of feedback was particularly beneficial for students who were struggling during the learning phase. These promising results regarding the benefit of worked examples as a way of delivering feedback are limited in that the present study did not systematically manipulate and compare various types of feedback. This leaves open the possibility that alternative types of feedback or support during learning may be more effective than studying worked examples.

General Discussion

Three experiments investigated the effectiveness of retrieval practice for learning complex and interrelated materials. In Experiments 1 and 2 retrieval practice required students to recall the procedural steps that are needed to solve Poisson probability problems. In Experiment 1, when the final assessment required students to recall the procedure, a retrieval practice effect was demonstrated. However, retrieval practice afforded no learning benefits in Experiment 2 when the assessment was a problem-solving test. Put simply, retrieval practice produced a large benefit when the assessment imposed minimal working memory demands, but this benefit disappeared when students were given a problem-solving test that induced a greater burden on working memory. This pattern of data appears to be consistent with cognitive load theory and its concept of element interactivity, but the transfer-appropriate processing framework provides an alternative explanation.

According to the idea of transfer-appropriate processing, optimal test performance depends on the compatibility between the cognitive processing that occurs during learning and the cognitive processing that is required by the test. Consistent with this idea, a pronounced retrieval practice effect was observed in Experiment 1 where there was high compatibility between the processing required during the retrieval practice condition and the recall test. This learning-test processing compatibility for the retrieval practice condition was eliminated in Experiment 2 because the assessment switched to a problem-solving test. As a result, neither learning activity in Experiment 2 uniquely afforded processing during learning that was compatible with the cognitive processing required by the test, which could explain the absence of a retrieval practice effect.

In order to evaluate these competing explanations, Experiment 3 pitted the predictions from cognitive load theory against the predictions from transfer-appropriate processing by having students learn to solve Poisson probability problems by studying worked examples or by solving practice problems. According to cognitive load theory and its concept of element interactivity, retrieval practice is an ineffective learning activity because it overburdens students limited working memory, which is already experiencing cognitive load due to the complexity of the materials. (van Gog et al., 2015). Thus, cognitive load theory predicts that studying worked examples would lead to better test performance compared to retrieval-based learning activities. Alternatively, the transfer-appropriate processing account predicts that solving practice problems would afford the same cognitive processing that would be critical on a problem-solving test, and this learning-test compatibility would facilitate test performance. The caveat with this prediction is that students must be engaging in test compatible processing during learning, which would be demonstrated by students solving some of the practice problems correctly. If learners are unable to solve practice problems successfully then they would not be engaging in test compatible processing during learning and would not have an advantage over the worked example condition on the test.

The results from Experiment 3 clashed with the predictions from cognitive load theory because studying worked examples decreased self-reported levels of mental effort but failed to produce a learning benefit. In fact, the worked example condition had the lowest test performance for all question types. The results from Experiment 3 are consistent with the idea of transfer-appropriate processing because the problem solving with feedback condition outperformed the worked example condition on repeated and near transfer questions, which required a similar procedure (or identical in the case of repeated questions) to what students had practiced during learning. In other words, these questions required the same cognitive processing that students in the problem-solving conditions had the opportunity to engage in during learning. Far transfer questions, on the other hand, required students to use a subset or variation of the procedure requiring different mental processing than what had been practiced during learning. In this case, none of the learning activities uniquely afforded test compatible processing during learning, which could explain the equivalent performance across the learning activities for far transfer questions.

The lack of a benefit from studying worked examples is even more striking as this learning activity was rated as the least difficult and required the least amount of mental effort. These lower ratings of difficulty and mental effort align with the prediction from cognitive load theory that worked examples create less cognitive load than solving practice problems (van Gog et al., 2015). This reduction in cognitive load from studying worked examples is said to free up limited working memory resources that are critical when learning complex materials (Paas & van Gog, 2006). Despite a reduction in mental effort for the worked example condition during learning, studying worked examples produced the lowest levels of test performance and highest judgments of learning. Studying worked examples was also rated as less engaging and less interesting compared to the problem-solving conditions. By studying with an unengaging and uninteresting learning activity that produces overconfidence, students are at risk of terminating their studying prematurely. In this way, studying worked examples could lead to tragically ineffective study choices, resulting in even less learning than solving practice problems.

Overall, the present findings, including the small-scale, exploratory meta-analysis presented in the introduction, identified a limitation with the worked example literature. Namely, the worked example effect is typically demonstrated when compared against impoverished learning activities — learning by repetitive failure (e.g., van Gog & Kester, 2012). Learning by repetitive failure is a straw-man version of retrieval practice, and the limited benefits of throwing problems at unprepared students is a poor comparison condition to demonstrate the efficacy of a learning activity. In the medical field, new treatments must outperform the current best practice in order to be recommended (Kornell et al., 2012). Education research needs to adopt this approach to better evaluate the effectiveness of learning activities before strong conclusions are drawn.

Proponents of cognitive load theory could try to argue that problem-solving success during learning constitutes expertise and results in the expertise reversal effect, the finding that studying worked examples eventually becomes redundant and ineffective when learners become experts (see Kalyuga, 2007 for a review). This argument is problematic because typical demonstrations of expertise reversal involve multi-session experiments where students see far more than the four questions presented in Experiment 3 (Kalyuga et al., 2001, Kalyuga & Sweller, 2004). Furthermore, classifying the students in Experiment 3 as experts seems inappropriate because they entered the study with little to no prior knowledge, struggled on the practice problems, and did not achieve high levels of test performance. Students do not need to become experts before they can experience a retrieval practice effect, but they do need to achieve some retrieval success during learning in order to benefit from retrieval-based learning.

To conclude, the claim that the retrieval practice effect is eliminated for complex materials is extremely harmful for teachers and students and is unsupported by both the extant literature and the present findings. Retrieval-based learning activities have been consistently found to be one of the best, if not the best learning activity that has been discovered, and all learners should be strongly encouraged to use retrieval practice when studying.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*(3), 849-858.

Chen, O., Kalyuga, S., & Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction, 45*, 20-30.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Darabi, A. A., Nelson, D. W., & Palanki, S. (2007). Acquisition of troubleshooting skills in a computer simulation: Worked example vs. conventional problem-solving instructional strategies. *Computers in Human Behavior*, *23*(4), 1809-1819.

Hanham, J., Leahy, W., & Sweller, J. (2017). Cognitive load theory, element interactivity, and the testing and reverse testing effects. *Applied Cognitive Psychology, 31*(3), 265-280.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational psychology review, 19*(4), 509-539.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402-410.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, *38*(1), 23–31.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579-588.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*(3), 558-568.

Kang, S. H. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition, 38*(8), 1009-1017.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (J. H. Byrne, Series Ed.) (pp. 487-514). Oxford: Academic Press.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27*(2), 317-326.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation, Vol. 61* (pp. 237-284). San Diego, CA: Elsevier Academic Press.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67*(1), 17-29.

Kelly, J. W., Carpenter, S. K., & Sjolund, L. A. (2015). Retrieval enhances route knowledge acquisition, but only when movement errors are prevented. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(5), 1540-1547.

Knouse, L. E., Rawson, K. A., Vaughn, K. E., & Dunlosky, J. (2016). Does testing improve learning for college students with attention-deficit/hyperactivity disorder? *Clinical Psychological Science*, *4*(1), 136-143.

Kornell, N., Rabelo, V. C., & Klein, P. J. (2012). Tests enhance learning—Compared to what? *Journal of Applied Research in Memory and Cognition*, *1*, 257-259.

Larsen, D. P., Butler, A. C., Lawson, A. L., & Roediger, H. L. (2013). The importance of seeing the patient: Test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Advances in Health Sciences Education*, *18*(3), 409-425.

Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, *27*(2), 291-304.

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science? *Journal of Applied Research in Memory and Cognition*, *3*(3), 171-176.

Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory*, *20*(8), 899-906.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*(3), 360-372.

McDermott, K. B., Agarwal, P. K., D'antonio, L., Roediger III, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*(1), 3-21.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519-533.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, *84*(4), 429-434.

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87-91.

Rawson, K. A. (2015). The status of the testing effect for complex materials: still a winner. *Educational Psychology Review*, *27*(2), 327–331.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory and research in educational psychology* (pp. 91–108). New York: Cambridge University Press.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*(9), 1043-1056.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1463.

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*(7), 784-802.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6), 1712–1725.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2), 165-169.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29-47). New York, NY, US: Cambridge University Press.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem Solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296

Van den Berge, K., van Gog, T., Mamede, S., Schmidt, H. G., van Saase, J. L. C. M., & Rikers, R. M. J. P. (2013). Acquisition of visual perceptual skills from worked examples: learning to interpret electrocardiograms (ECGs). *Interactive Learning Environments*, *21*(3), 263-272.

Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, *36*(8), 1532-1541.

Van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*(2), 265-289.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212-218.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155–174.

Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*(2), 247-264.

Yeo, D. J., & Fazio, L. K. (2019). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *Journal of Educational Psychology*, *111*(1), 73.

Zamary, A., & Rawson, K. A. (2018). Are provided examples or faded examples more effective for declarative concept learning? *Educational Psychology Review*, *30*(3), 1167–1197.

Appendix A

The Poisson Distribution

A Bernoulli trial is one in which the outcome is either a success or a failure. An example would be flipping a coin where heads is considered a "success" and tails is considered a "failure." Often, over the course of a series of Bernoulli trials, the most important information is not which trials ended in success and which in failure, but rather, **how many** ended in success or failure. Let $\boldsymbol{X}$ denote the number of successes in $\boldsymbol{n}$ Bernoulli trials, when the probability of a success on any particular trial is $\boldsymbol{p}$. Then $\boldsymbol{X}$ is said to have a Binomial distribution, $\boldsymbol{X \sim \mathrm{B}\,(n, p)}$, and the probability of getting $\boldsymbol{x}$ successes in $\boldsymbol{n}$ trials is:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad , x = 0, 1, 2, \dots, n.$$

The equation above can sometimes get quite messy when $\boldsymbol{n}$ and $\boldsymbol{x}$ get large. For certain events that occur **singly**, **independently** and **randomly**, with the **probability** $\boldsymbol{p}$ of one **event occurring** within a small fixed interval of time (or space) is the **same** and **fairly low** at all points in time (or space), we can often use the Poisson distribution, $\boldsymbol{X \sim \mathrm{P_o}(\lambda)}$, as a replacement for the Binomial distribution to model the frequency of the occurrence of the events. We can replace the Binomial equation with the Poisson equation:

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad , x = 0, 1, 2, \dots, n, and\ \lambda > 0$$

where $\boldsymbol{e} \approx 2.718$, and where $\boldsymbol{\lambda}$ is the expected value (that is, the **average** or **mean** value) of the random variable $\boldsymbol{X}$. This equation is much easier to calculate for the various values of $\boldsymbol{X}$ than the Binomial equation.

The distribution has a mean number (or expected number) of occurrences, $\boldsymbol{\lambda}$, in a given time (or space) that is proportional to the time (or space) interval. For example, if $\boldsymbol{\lambda}$ is the mean number of phone calls received in a 1 minute interval, then the mean number of phone calls received in a 2 minute interval will be equal to $2\boldsymbol{\lambda}$.

If $X \sim \mathrm{P_o}(\lambda_1)$ and $Y \sim \mathrm{P_o}(\lambda_2)$, where X and Y are **independent**, and $W = X + Y$ (sum of Poisson random variables), then W is also a Poisson random variable, $W \sim \mathrm{P_o}(\lambda_1 + \lambda_2)$.

***

Instructions on how to use the Texas Instruments TI-30XS MultiView calculator

To compute the value of a number raised to a power n:

For example, to find $10^{-3.2}$, press: **1 0** $\boxed{\wedge}$ $\boxed{(-)}$ **3 . 2** $\boxed{\text{enter}}$

To compute the factorial of integer n:

For example, to find 7!, press: **7** $\boxed{\text{prb}}$ **3** $\boxed{\text{enter}}$

Appendix B

The Procedural Steps that were Studied During Experiments 1 and 2

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Step 4: Determine the boundaries of X

Step 5: Expand the formula for each value contained within the boundaries of X

Step 6: Calculate the probability

Appendix C

The Worked Examples for the Learning Phase of Experiment 3

Worked Example 1

Suppose that the arrival and departure of airplanes at a domestic airport follow two independent Poisson distributions. In a one-hour period, it is expected on average that there are 4 arrivals and 3 departures. Find the probability that, in a randomly selected two-hour period, the airport handles 10 or more, but less than 13 arrivals and departures.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of arrivals in a two-hour period
1 hour $\rightarrow$ 4 arrivals
2 hours $\rightarrow 4 \times 2 = 8$ arrivals
So, $V_1 \sim P_o(8)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of departures in a two-hour period
1 hour $\rightarrow$ 3 departures
2 hours $\rightarrow 3 \times 2 = 6$ departures
So, $V_2 \sim P_o(6)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of arrivals and departures in a two-hour period
$T \sim P_o(8 + 6)$
$T \sim P_o(14)$

Step 4: Determine the boundaries of X

10 or more and less than 13
$(10 \leq X < 13)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(10 \leq X < 13) = P(X = 10) + P(X = 11) + P(X = 12)$

$$= \frac{e^{-14} \cdot 14^{10}}{10!} + \frac{e^{-14} \cdot 14^{11}}{11!} + \frac{e^{-14} \cdot 14^{12}}{12!}$$

Step 6: Calculate the probability

= .249

Worked Example 2

Suppose that the number of calls received by a hospital and a fire station follow two independent Poisson distributions. In a one-week period, it is expected on average that a hospital receives 11 calls and a fire station receives 5 calls. Find the probability that, in a randomly selected three-week period, the hospital and the fire station receive a total of more than 52, but less than 56 calls.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of calls received by the hospital in a three-week period

1 week $\rightarrow$ 11 calls

3 weeks $\rightarrow$ 11 $\times$ 3 = 33 calls

So, $V_1 \sim P_o(33)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of calls received by the fire station in a three-week period

1 week $\rightarrow$ 5 calls

3 weeks $\rightarrow$ 5 $\times$ 3 = 15 calls

So, $V_2 \sim P_o(15)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of calls received by the hospital and fire station in a three-week period

$T \sim P_o(33 + 15)$

$T \sim P_o(48)$

Step 4: Determine the boundaries of X

More than 52 and less than 56

$(52 < X < 56)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(52 < T < 56) = P(T = 53) + P(T = 54) + P(T = 55)$

$$= \frac{e^{-48} \cdot 48^{53}}{53!} + \frac{e^{-48} \cdot 48^{54}}{54!} + \frac{e^{-48} \cdot 48^{55}}{55!}$$

Step 6: Calculate the probability

= .113

Worked Example 3

Suppose that, in a restaurant, the number of cups and saucers being broken each day while washing follow two independent Poisson distributions. In one day, it is expected on average that there are 2 cups and 1 saucer broken. Find the probability that, in a randomly selected seven-day period, the total number of cups broken and saucers broken is at least 18 but no more than 20.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of cups broken in a seven-day period
1 day → 2 cups
7 days → 2 × 7 = 14 cups
So, $V_1 \sim P_o(14)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of saucers broken in a seven-day period
1 day → 1 saucer
7 days → 1 × 7 = 7 saucers
So, $V_2 \sim P_o(7)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of cups and saucers broken in a seven-day period
$T \sim P_o(14 + 7)$
$T \sim P_o(21)$

Step 4: Determine the boundaries of X

At least 18 but no more than 20
$(18 \leq X \leq 20)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(18 \leq T \leq 20) = P(T = 18) + P(T = 19) + P(T = 20)$

$$= \frac{e^{-21} \cdot 21^{18}}{18!} + \frac{e^{-21} \cdot 21^{19}}{19!} + \frac{e^{-21} \cdot 21^{20}}{20!}$$

Step 6: Calculate the probability

= .244

Worked Example 4

Suppose that the number of unsolicited text messages and phone calls received by a mobile line subscriber follow two independent Poisson distributions. In a one-week period, it is expected on average that there are 5 unsolicited text messages and 3 unsolicited phone calls. Find the probability that, in a randomly selected four-week period, the subscriber receives more than 37 but at most 40 unsolicited text messages or phone calls.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of unsolicited text messages in a four-week period
1 week $\rightarrow$ 5 messages
4 weeks $\rightarrow$ $5 \times 4 = 20$ messages
So, $V_1 \sim P_o(20)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of unsolicited phone calls in a four-week period
4 weeks $\rightarrow$ $3 \times 4 = 12$ calls
So, $V_2 \sim P_o(12)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of unsolicited text messages and phone calls in a four-week period
$T \sim P_o(20 + 12)$
$T \sim P_o(32)$

Step 4: Determine the boundaries of X

More than 37 but at most 40
$(37 < X \leq 40)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(37 < T \leq 40) = P(T = 38) + P(T = 39) + P(T = 40)$

$$= \frac{e^{-32}. 32^{38}}{38!} + \frac{e^{-32}. 32^{39}}{39!} + \frac{e^{-32}. 32^{40}}{40!}$$

Step 6: Calculate the probability

= .094

Appendix D

Practice Problems for Learning Phase of Experiment 3

Practice problems have the same question prompt as the corresponding worked example and were presented on individual pages.

Practice Problem 2 (Please label each step and show all of your work)

Suppose that the number of calls received by a hospital and a fire station follow two independent Poisson distributions. In a one-week period, it is expected on average that a hospital receives 11 calls and a fire station receives 5 calls. Find the probability that, in a randomly selected three-week period, the hospital and the fire station receive a total of more than 52, but less than 56 calls.

Practice Problem 3 (Please label each step and show all of your work)

Suppose that, in a restaurant, the number of cups and saucers being broken each day while washing follow two independent Poisson distributions. In one day, it is expected on average that there are 2 cups and 1 saucer broken. Find the probability that, in a randomly selected seven-day period, the total number of cups broken and saucers broken is at least 18 but no more than 20.

Practice Problem 4 (Please label each step and show all of your work)

Suppose that the number of unsolicited text messages and phone calls received by a mobile line subscriber follow two independent Poisson distributions. In a one-week period, it is expected on average that there are 5 unsolicited text messages and 3 unsolicited phone calls. Find the probability that, in a randomly selected four-week period, the subscriber receives more than 37 but at most 40 unsolicited text messages or phone calls.

Appendix E

Test Problems for Experiments 2 and 3

Students were given a test packet with one question prompt presented per page. Questions 1 and 2 are near transfer questions, questions 3-8 are far transfer questions, and questions 9-11 are repeated questions from the learning phase (i.e., practice problems 2, 3, 4). Only Experiment 3 included test questions 9, 10, and 11. The answer to each problem is provided in worked example format for convenience, but students only saw the question prompt when tested. Below each question number is the type of question, which was not shown to students.

Test Problem 1 (Please label each step and show all of your work)
(Near transfer: two events and larger required interval)
A car salesperson sells, on average, 3 new cars and 2 used cars in two weeks. The number of new cars she sells is independent of the number of old cars she sells, and they each follow independent Poisson distributions. Find the probability that she sells at least 5 but at most 7 cars in a randomly chosen four-week period.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1be the number of new cars she sells in a four-week period
2 weeks $\rightarrow$ 3 new cars
4 weeks $\rightarrow 3 \times 2 = 6$ new cars
So, $V_1 \sim P_o(6)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of used cars she sells in a four-week period
2 weeks $\rightarrow$ 2 used cars
4 weeks $\rightarrow 2 \times 2 = 4$ used cars
So, $V_2 \sim P_o(4)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of new and used cars she sells in a four-week period
$T \sim P_o(6 + 4)$
$T \sim P_o(10)$

Step 4: Determine the boundaries of X

$(5 \leq X \leq 7)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(10 \leq T < 13) = P(T = 5) + P(T = 6) + P(T = 7)$
$$= \frac{e^{-10} \cdot 10^5}{5!} + \frac{e^{-10} \cdot 10^6}{6!} + \frac{e^{-10} \cdot 10^7}{7!}$$

Step 6: Calculate the probability

$= .191$

Test Problem 2 (Please label each step and show all of your work)
(Near transfer: two events and larger required interval)

A university has two departments and each department records the number of employees absent through illness each day. Each employee absent for a day represents one "day of absence'. So, one employee absent for 2 days contributes 2 days of absence, and 6 employees absent on 1 day contribute 6 days of absence. Over a long period of time it is found that the average numbers absent for a day are 1.6 for Psychology Department and 2.2 for the Biomedical Department. Suppose that the absences in the two departments are independent of each other, and they each follow independent Poisson distributions, find the probability that, in a randomly chosen 5-day period, the total number of days of absence in the two departments is more than 21 but less than 24.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of days of absence in a 5-day period for the Psychology Department

1 day $\rightarrow$ 1.6 absences

5 days $\rightarrow$ $1.6 \times 5 = 8$ absences

So, $V_1 \sim P_o(8)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of days of absence in a 5-day period for the Biomedical Department.

1 day $\rightarrow$ 2.2 absences

5 days $\rightarrow$ $2.2 \times 5 = 11$ absences

So, $V_2 \sim P_o(11)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of days of absence in a 5-day period for the Psychology and Biomedical Departments

$T \sim P_o(8 + 11)$

$T \sim P_o(19)$

Step 4: Determine the boundaries of X

$(21< X < 24)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(21< X < 24)= P(T = 22) + P(T = 23)$

$$= \frac{e^{-19} \cdot 19^{22}}{22!} + \frac{e^{-19} \cdot 19^{23}}{23!}$$

Step 6: Calculate the probability

$= .124$

Test Problem 3 (Please label each step and show all of your work)
(Far transfer: single event and smaller required interval)

At a newly opened bistro, the number of orders for clam chowder received in a randomly chosen one-hour period follows a Poisson distribution with mean 4.6. Find the probability that there are less than 2 orders received in a randomly chosen 30-minute interval.

Step 1: For the given time period, let Variable 1 (V_1) equal the average and convert it to the randomly selected period

Let V_1 be the number of orders for clam chowder received in a 30-minute interval

1 hour $\rightarrow$ 4.6 orders

30 minutes $\rightarrow 4.6 \div 2 = 2.3$ orders

So, $V_1 \sim P_o(2.3)$

Step 2: Determine the boundaries of X

$(0 \leq X < 2)$

Step 3: Expand the formula for each value contained within the boundaries of X

$P(0 \leq X < 2) = P(T = 0) + P(T = 1)$

$$= \frac{e^{-2.3} \cdot 2.3^0}{0!} + \frac{e^{-2.3} \cdot 2.3^1}{1!}$$

Step 4: Calculate the probability

= .331

Test Problem 4 (Please label each step and show all of your work)
(Far transfer: single event and smaller required interval)

The occurrences of floods per year at a particular residential area in Swiftville follow a Poisson distribution with mean 5. Find the probability that in a randomly chosen 3-month period, this particular residential area is flooded at most twice.

Step 1: For the given time period, let Variable 1 (V_1) equal the average and convert it to the randomly selected period

Let V_1 be the number of floods in a 3-month period
12 months $\rightarrow$ 5 floods
3 months $\rightarrow$ $5 \div 4 = 1.25$ floods
So, $V_1 \sim P_o(1.25)$

Step 2: Determine the boundaries of X

$(0 \leq X \leq 2)$

Step 3: Expand the formula for each value contained within the boundaries of X

$P(0 \leq F \leq 2) = P(F = 0) + P(F = 1) + P(F = 2)$

$$= \frac{e^{-1.25} \cdot 1.25^0}{0!} + \frac{e^{-1.25} \cdot 1.25^1}{1!} + \frac{e^{-1.25} \cdot 1.25^2}{2!}$$

Step 4: Calculate the probability

= .868

Test Problem 5 (Please label each step and show all of your work)
(Far transfer: two events and different given intervals)

The two most common types of disciplinary offenses in a particular boys school in England is keeping long hair and failure to wear the school badge. Assuming that each school week consists of five school days and each school month consists of 20 school days, the mean number of disciplinary offenses recorded per day involving long hair is 1.35, and the mean number of disciplinary offenses recorded per school month involving failure to wear the school badge is 5. The number of cases for each disciplinary offense is assumed to have an independent Poisson distribution. Find the probability that more than 8 and less than 11 cases of disciplinary offenses are recorded in a randomly chosen week.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of disciplinary offences recorded for long hair per week
1 day $\rightarrow$ 1.35 cases
5 days $\rightarrow 1.35 \times 5 = 6.75$ cases
So, $V_1 \sim P_o(6.75)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of disciplinary offences recorded for failure to wear the school badge per week.
20 days $\rightarrow$ 5 cases
5 day $\rightarrow \frac{5}{20} \times 5 = 1.25$ cases
So, $V_2 \sim P_o(1.25)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of cases of disciplinary offences recorded in a week
$T \sim P_o(6.75 + 1.25)$
$T \sim P_o(8)$

Step 4: Determine the boundaries of X.

$(8 < T < 11)$

Step 5: Expand the formula for each value contained within the boundaries of X.

$P(8 < T < 11) = P(T = 9) + P(T = 10)$

$$= \frac{e^{-8} \cdot 8^9}{9!} + \frac{e^{-8} \cdot 8^{10}}{10!}$$

Step 6: Calculate the probability

= .223

Test Problem 6 (Please label each step and show all of your work)
(Far transfer: two events and different given intervals)

Vehicles travelling towards the city center pass a toll booth at an average rate of 4.5 per 30-minute interval. Vehicles travelling away from the town center pass the same toll booth at an average rate of 2 per 20-minute interval. Suppose that they follow independent Poisson distributions, find the probability that there will be more than 18 but no more than 21 vehicles passing the toll booth between 8 am to 9 am on a randomly chosen day.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of vehicles travelling towards the city center in 1 hour
30 minutes $\rightarrow$ 4.5 vehicles
1 hour $\rightarrow 4.5 \times 2 = 9$ vehicles
So, $V_1 \sim P_o(9)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of vehicles travelling away from the city center in 1 hour
20 minutes $\rightarrow$ 2 vehicles
1 hour $\rightarrow 2 \times 3 = 6$ vehicles
So, $V_2 \sim P_o(6)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of vehicles travelling towards and away from the city center in 1 hour
$T \sim P_o(9 + 6)$
$T \sim P_o(15)$

Step 4: Determine the boundaries of X

$(18 < T \leq 21)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(18 < T \leq 21) = P(T = 19) + P(T = 20) + P(T = 21)$

$$= \frac{e^{-15} \cdot 15^{19}}{19!} + \frac{e^{-15} \cdot 15^{20}}{20!} + \frac{e^{-15} \cdot 15^{21}}{21!}$$

Step 6: Calculate the probability

= .127

Test Problem 7 (Please label each step and show all of your work)
(Far transfer: two events with the same mean, but with different intervals within the same "space")

In the production of cellphone screen protectors, scratches occur at random and independently, and they follow a Poisson distribution with a mean of 0.15 scratches per screen protector. In a quality control inspection, 100 screen protectors produced by manufacturer A and 200 screen protectors produced by manufacturer B were selected randomly. Find the probability that there are more than 51 but no more than 54 scratches in a randomly selected quality control inspection.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of scratches found on 100 screen protectors produced by manufacturer A in an inspection
1 screen protectors $\rightarrow$ 0.15 scratches
100 screen protectors $\rightarrow 0.15 \times 100 = 15$ scratches
So, $V_1 \sim P_o(15)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of scratches found on 200 screen protectors produced by manufacturer B in an inspection.
1 screen protectors $\rightarrow$ 0.15 scratches
200 screen protectors $\rightarrow 0.15 \times 200 = 30$ scratches
So, $V_2 \sim P_o(30)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of scratches found on the 300 screen protectors produced by manufacturers A and B in a randomly selected inspection
$T \sim P_o(15 + 30)$ i.e., $T \sim P_o(45)$

Step 4: Determine the boundaries of X

$(51 < T \leq 54)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(51 < T \leq 54) = P(T = 52) + P(T = 53) + P(T = 54)$

$$= \frac{e^{-45} \cdot 45^{52}}{52!} + \frac{e^{-45} \cdot 45^{53}}{53!} + \frac{e^{-45} \cdot 45^{54}}{54!}$$

Step 6: Calculate the probability

= .084

Test Problem 8 (Please label each step and show all of your work)
(Far transfer: two events with the same mean, but with different intervals within the same "space")

Two identical racing robots are being tested on a circuit. For each robot, the number of mechanical breakdowns follows a Poisson distribution with a mean of 2 breakdowns in 100 laps. Robot X does 40 laps and Robot Y does 60 laps in a test respectively. Assuming that the breakdowns are attended to, and the robots continue on the circuit, find the probability that there will be at least 1 but less than 3 breakdowns altogether during a randomly selected test.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of breakdowns for Robot X in 40 laps
100 laps $\rightarrow$ 2 breakdowns
40 laps$\rightarrow \frac{2}{100} \times 40$ = 0.8 breakdowns
So, $V_1 \sim P_o(0.8)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of breakdowns for Robot Y in 60 laps
100 laps $\rightarrow$ 2 breakdowns
60 laps$\rightarrow \frac{2}{100} \times 60$ = 1.2 breakdowns
So, $V_2 \sim P_o(1.2)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of breakdowns altogether during a randomly selected test
$T \sim P_o(0.8 + 1.2)$
$T \sim P_o(2)$

Step 4: Determine the boundaries of X

$(1 \leq T < 3)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(1 \leq T < 3) = P(T = 1) + P(T = 2)$

$$= \frac{e^{-2} \cdot 2^1}{1!} + \frac{e^{-2} \cdot 2^2}{2!}$$

Step 6: Calculate the probability

= .541

Test Problem 9 (Please label each step and show all of your work)
(Repeated: practice problem 2)

Suppose that the number of calls received by a hospital and a fire station follow two independent Poisson distributions. In a one-week period, it is expected on average that a hospital receives 11 calls and a fire station receives 5 calls. Find the probability that, in a randomly selected three-week period, the hospital and the fire station receive a total of more than 52, but less than 56 calls.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of calls received by the hospital in a three-week period
1 week $\rightarrow$ 11 calls
3 weeks $\rightarrow$ $11 \times 3 = 33$ calls
So, $V_1 \sim P_o(33)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of calls received by the fire station in a three-week period
1 week $\rightarrow$ 5 calls
3 weeks $\rightarrow$ $5 \times 3 = 15$ calls
So, $V_2 \sim P_o(15)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of calls received by the hospital and fire station in a three-week period
$T \sim P_o(33 + 15)$
$T \sim P_o(48)$

Step 4: Determine the boundaries of X

More than 52 and less than 56
$(52 < X < 56)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(52 < T < 56) = P(T = 53) + P(T = 54) + P(T = 55)$

$$= \frac{e^{-48} \cdot 48^{53}}{53!} + \frac{e^{-48} \cdot 48^{54}}{54!} + \frac{e^{-48} \cdot 48^{55}}{55!}$$

Step 6: Calculate the probability

$= .113$

Test Problem 10 (Please label each step and show all of your work)
(Repeated: practice problem 3)

Suppose that, in a restaurant, the number of cups and saucers being broken each day while washing follow two independent Poisson distributions. In one day, it is expected on average that there are 2 cups and 1 saucer broken. Find the probability that, in a randomly selected seven-day period, the total number of cups broken and saucers broken is at least 18 but no more than 20.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of cups broken in a seven-day period
1 day $\rightarrow$ 2 cups
7 days $\rightarrow 2 \times 7 = 14$ cups
So, $V_1 \sim P_o(14)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of saucers broken in a seven-day period
1 day $\rightarrow$ 1 saucer
7 days $\rightarrow 1 \times 7 = 7$ saucers
So, $V_2 \sim P_o(7)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let λ be the total number of cups and saucers broken in a seven-day period
$T \sim P_o(14 + 7)$
$T \sim P_o(21)$

Step 4: Determine the boundaries of X

At least 18 but no more than 20
$(18 \leq X \leq 20)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(18 \leq T \leq 20) = P(T = 18) + P(T = 19) + P(T = 20)$

$$= \frac{e^{-21} \cdot 21^{18}}{18!} + \frac{e^{-21} \cdot 21^{19}}{19!} + \frac{e^{-21} \cdot 21^{20}}{20!}$$

Step 6: Calculate the probability

= .244

Test Problem 11 (Please label each step and show all of your work)
(Repeated: practice problem 4)

Suppose that the number of unsolicited text messages and phone calls received by a mobile line subscriber follow two independent Poisson distributions. In a one-week period, it is expected on average that there are 5 unsolicited text messages and 3 unsolicited phone calls. Find the probability that, in a randomly selected four-week period, the subscriber receives more than 37 but at most 40 unsolicited text messages or phone calls.

Step 1: For the given time period, let Variable 1 (V_1) equal one of the averages and convert it to the randomly selected period

Let V_1 be the number of unsolicited text messages in a four-week period

1 week $\rightarrow$ 5 messages

4 weeks $\rightarrow$ $5 \times 4 = 20$ messages

So, $V_1 \sim P_o(20)$

Step 2: For the given time period, let Variable 2 (V_2) equal the other average and convert it to the randomly selected period

Let V_2 be the number of unsolicited phone calls in a four-week period

4 weeks $\rightarrow$ $3 \times 4 = 12$ calls

So, $V_2 \sim P_o(12)$

Step 3: Add Steps 1 and 2 together to find the expected number of occurrences (λ)

Let T be the total number of unsolicited text messages and phone calls in a four-week period

$T \sim P_o(20 + 12)$

$T \sim P_o(32)$

Step 4: Determine the boundaries of X

More than 37 but at most 40

$(37 < X \leq 40)$

Step 5: Expand the formula for each value contained within the boundaries of X

$P(37 < T \leq 40) = P(T = 38) + P(T = 39) + P(T = 40)$

$$= \frac{e^{-32} \cdot 32^{38}}{38!} + \frac{e^{-32} \cdot 32^{39}}{39!} + \frac{e^{-32} \cdot 32^{40}}{40!}$$

Step 6: Calculate the probability

= .094