Spring 2022

Improving Student Outcomes in Introductory Formal Logic

Bryce Rosenwald

Improving Student Outcomes in Introductory Formal Logic

Bryce Rosenwald
Honors College, Western Washington University

Advisor: Ryan Wasserman, Ph.D.
Department of Philosophy, Western Washington University

June 2022

Abstract

Across a variety of educational settings, undergraduate introductory courses in formal logic tend to have high failure rates. In this paper, I explore practical, evidence-based steps that logic instructors can take to improve student outcomes. The topics covered are small class sizes, problem-based learning, clicker questions, group activities, and spaced practice. The effect of small class size is moderated by many variables, including each instructor's unique characteristics and the pedagogical techniques used in large vs. small classrooms. Problem-based learning and clicker questions are determined to be excellent techniques for introducing content and furthering understanding of content, respectively. Small groups can enhance both types of activity. Finally, incorporating spaced practice into homework assignments reliably improves retention of material. Possible challenges to incorporating these techniques to the logic classroom are described, and potential solutions are discussed.

Acknowledgements

I am deeply grateful to Ryan Wasserman, the advisor of this project, for his generous support and guidance at every step of the process. I would also like to thank Christian Lee, Dan Howard-Snyder, Ian Schnee, Ross Colebrook, Russ Payne, Jim Graham, Reece Brown, and Kaden Dagher, all of whom contributed to or supported this project in some way.

Improving Student Outcomes in Introductory Formal Logic

I. Introduction

The department of philosophy at Western Washington University offers an introductory course in formal logic. This course, Philosophy 102, is offered every quarter, is open to all students, and fulfills a Quantitative and Symbolic Reasoning (QSR) general education requirement. Students in this course learn to symbolize English sentences and arguments in propositional symbolic logic, evaluate arguments using truth tables, and demonstrate validity by completing proofs using a system of inference rules.

Although many students do well in Philosophy 102, the proportion of students who fail is consistently high. In 2016, 16.6 percent of students failed the course, making it the most failed course at the university ("When All Else Fails," 2017). The second-most failed course was Math 112 (Functions and Algebraic Methods), with a failure rate of 15.9 percent, and the third-most failed was Math 115 (Precalculus II), with a failure rate of 12.6 percent. Like Philosophy 102, Math 112 and Math 115 are both introductory courses that fulfill a QSR requirement, suggesting that the similar failure rates among these courses reflect common difficulties teaching courses with these attributes. Unsurprisingly, outcomes in Philosophy 102 have been especially poor during the COVID-19 pandemic.

Student success in formal logic is a challenge for many philosophy departments at a variety of institutions. Bellevue College, a large public community college in Bellevue, Washington, offers an introductory course in propositional and predicate logic. From fall 2016 to spring 2021, 36.3 percent of students in this course received a C- or lower (W. Payne, personal communication, February 11, 2022). At Medgar Evers College, a mid-size public college in Brooklyn primarily serving graduates of the New York City public school system, approximately 15 to 20 percent of students in introductory logic either drop or fail the course (R. Colebrook, personal communication, June 3, 2022). Regarding why teaching logic proves such a challenge, Pearlman (2019) notes that students must learn abstract conceptual content, become familiar with an artificial logical language, and acquire the skills to solve relevant sorts of problems. Thus, logic combines the most difficult elements of philosophy, foreign language, and mathematics into a single course.

Without some sort of intervention, it is likely that high failure rates in logic courses will persist. Failure rates in Philosophy 102 were rising even before the pandemic, with the rate increasing from 11.2 percent in 2009 to 16.6 percent in 2016 ("When All Else Fails," 2017). Furthermore, standardized test results suggest that students are leaving high school extremely ill-prepared for college-level math courses. Only 24 percent of Washington eleventh graders met standards on statewide math exams taken in fall 2021 (Bazzaz, 2022). More broadly, university faculty across the country report that in the wake of the pandemic, students are exhausted, apathetic, and disengaged, and it is unclear how long this trend will last (McMurtrie, 2022).

In this paper, I explore modest, evidence-based steps that logic instructors can take to improve student outcomes. The five topics covered are small class sizes, problem-based learning, clicker questions, group work, and spaced practice. I discuss the theoretical and empirical basis of each topic and, when applicable, provide examples of how each be implemented in the logic classroom. With the exception of class size, I do not review changes that would need to be made on the departmental or institutional level, such as providing out-of-class tutoring or study skills sessions.

II. Class Size

The effect of class size on student outcomes has been studied extensively. Although providing small class sizes is more expensive and resource-intensive, proponents argue that they benefit students by promoting active learning and student engagement and allowing instructors to provide more thorough feedback (Cuseo, 2007). Could student outcomes in formal logic courses be improved by reducing class size?

Two limitations of the literature on class size should be noted. First, although a great deal of research on the effects of class size has been conducted, most of this research focuses on K-12 students. Furthermore, that research suggests that in general, the beneficial effects of small class sizes diminish as students age (Biddle & Berliner, 2002; Shipman & Duch, 2001). Perhaps older students are more capable of managing their own learning and independently overcoming obstacles, so they may benefit less from the additional teacher attention afforded by smaller class sizes. Thus, research which shows benefits of small class sizes in K-12 education cannot be generalized to university courses, and we should expect any benefits of small class sizes in the university setting to be small. Second, most research on class size among older students is correlational analysis of actual student grades rather than the results of controlled experiments. Although this improves the external validity of the research, it limits our ability to establish unambiguous causal relationships between class size and student outcomes. It may be especially difficult to conduct a controlled experiment on class size which accurately mimics the conditions of a college course owning to the time and resources required, as well as the onerous obligations which would be imposed on the participants in such a study.

Jarvis (2007) conducted a multiple regression analysis on the final exam scores of 1,984 students in undergraduate first-semester calculus. The sample included both small classes of 20 to 35 students and large classes of 150 to 240 students. These classes were taught by 27 instructors, 4 of whom taught at least one small section and one large section. In the context of other predictors, including ACT math scores and pretest scores, class size did not emerge as a statistically significant predictor of final exam score. However, in the same analysis, the teacher-class size interaction was a statistically significant predictor of final exam score. In other words, across all students, class size was not correlated with final exam score, but among the students of those instructors who taught both large and small sections, class size may have been correlated with final exam score. Additional analyses demonstrated that, for three of those instructors, students performed better in the smaller classes, but for one professor, students performed better in the large class. Furthermore, the best instructors of large sections had a greater statistical effect on final grades than most instructors of smalls sections. The data from this study are not consistent with the conclusion that reducing class size is enough to improve student outcomes. Instead, optimal class size appears to be a function of the instructor's unique capabilities and strengths – some may be more effective teaching small classes while others may be more effective teaching large classes, and a large class with an excellent instructor can produce outcomes just as good as or better than a small class with an average instructor.

In a quasi-experiment conducted by Olson et al. (2011), students enrolled in one four sections of precalculus, each of which met four days a week. Unbeknownst to the students at the time of enrollment, the sections employed four different teaching approaches: a traditional lecture format in a large class (138 students), traditional lecture format in a small class (40 students), a modified lecture format in which students worked in small groups (61 students), and a modified lecture format with a problem-based emphasis (75 students). In the small-group format, students spent one day each week collaborating on a set of practice problems. In the problem-based emphasis format, practice problems were interspersed throughout each lecture. These problems prompted students to apply content to new situations and contexts. All students complete five exams during the course, standardized across section. The exam scores of the students in the problem-based emphasis section were statistically significantly higher than the scores from the other three sections, while there was no statistically significant difference between the scores of the large lecture section, the small lecture section, and the small-groups section. These results suggest that reducing class size has a minimal or no effect on student outcomes if teaching approach is not changed. Furthermore, the success of the 75 students in the problem-based emphasis section relative to the 40 students in the small lecture section demonstrates that a small class size is not necessary for student success. Instead, a relatively large class with a more effective format can produce outcomes as good or better than a relatively small class with a less effective format.

Determining the optimal class size for an undergraduate formal logic course is far more complex than "smaller is better." First, the unique characteristics of each instructor should be considered. Second, changing class size may produce no effect if the teaching approach and format remains constant. If so, then the benefit of smaller class sizes would come from the greater flexibility afforded to instructors to implement non-traditional teaching approaches. What, then, are the most effective teaching approaches, and do they require small classes?

III. Problem-Based Learning

A major meta-analysis of research on undergraduate science, math, engineering, and technology courses which measured a wide variety of student outcomes found that, in that context, active learning is reliably more effective than traditional lecturing (Freeman et al., 2014). "Active learning" is a broad category, and the research included in this meta-analysis featured a wide variety of manipulations. One active-learning technique which could be implemented in the logic classroom is problem-based learning (PBL). PBL can be understood as an alternative way to introduce content to students (Duch et al., 2001). In a traditional classroom, content is presented and explained to students, either via lecture or textbook, and they are then asked to apply and practice it. With PBL, students are presented not with content, but with problems. These problems are designed such that students can make some progress toward a solution using prior knowledge, common sense, and creativity. After working independently, students discuss their ideas and approaches. Finally, the instructor can summarize the most important insights from the discussion and provide a mini lecture to explicitly introduce content, if appropriate. The specifics of how PBL is implemented can vary drastically based on course goals, subject matter, and various practical constraints. For instance, students can approach the problems alone or in groups, and the amount of time spent on each problem can range from minutes to weeks (Duch, 2001; Shipman & Duch, 2001).

Problem-based learning emphasizes the role of specific cognitive activities in learning (Dolmans et al., 2005). Compared to traditional content presentation, PBL stimulates much more metacognition. In the process of working toward a solution, students will recognize the limits of what they already know. They may experiment, combine pieces of prior knowledge in new ways, and notice deeper conceptual connections within the content, thereby enhancing their understanding. PBL is also an opportunity to practice critical thinking, creativity, flexibility, and communication, skills which have benefits far beyond the logic classroom (Duch et al., 2001).

In a synthesis of eight meta-analyses comparing PBL to traditional classrooms, Strobel and van Barneveld (2009) found that PBL was superior for long-term retention of knowledge and skills and led to greater satisfaction among instructors and students. On the other hand, traditional classrooms led to higher performance on short-term retention tasks. PBL was developed in the field of medical education, so most published research on its efficacy, including that cited by Strobel and van Barneveld, concerns medicine and related fields (Yew & Goh, 2016). Given that medicine draws heavily on sciences including biology and chemistry, the research may be generalizable to other university-level technical or STEM fields. However, more research is still needed to determine the precise effect of PBL in mathematics and logic. In one promising study of high school algebra and trigonometry, Fletcher (2018) found that PBL improved student achievement and attitudes.

Dolmans et al. (2005) describe three challenges often encountered when implementing PBL. The first is that instructors often write problems which are too well-structured and do not challenge students. There are no simple rules for writing problems that effectively simulate active learning, and problems may need to be tested and revised to find the ideal degree of open-endedness. The authors suggest that drawing on real-world situations to craft problems can introduce the appropriate amount of messiness and extraneous information. The second challenge is that instructors are often too directive. An instructor who offers too much guidance can nullify any potential benefits of a well-crafted problem. One instructor suggested that, along with crafting problems, the most important skill in implementing PBL is overcoming the "instinct to teach" (I. Schnee, personal communication, April 14, 2022). The third challenge is that, when students approach problems in groups, these groups are sometimes dysfunctional. I will discuss some best practices for promoting effective student groups in section V.

PBL is a way to deliver content, as opposed to a way to practice skills or do something else. So, in introductory logic, it may be most useful during the beginning of the course, when foundational concepts and vocabulary (e.g., validity, formal validity, soundness) are introduced, and throughout the course whenever supplemental concepts (e.g., tautology, contingent statement, and contradiction) appear. Regarding the use of PBL in large classes, Shipman & Duch (2001) found that it remained effective in a large class of 120 students and a very large class of 240 students. However, they argue that instructors using PBL in large classes must introduce more structure to the class session, perhaps by limiting the time spent working on and discussing a problem to 15 minutes and incorporating more mini-lectures. In this format, the scope of problems must be fairly modest, but they should still be open-ended and instructors should still allow students to explore the problems without unnecessary guidance. In summary, problem-based learning can be an effective way to introduce new concepts and content to logic students, even in large classes of a few hundred students. Sample problems designed to facilitate learning of some key topics from introductory logic can be found in Appendix A.

IV. Clicker Activities

Clicker systems allow an instructor to present a practice problem or question to students, which is then answered via an electronic device. The instructor has immediate access to student answers, which can be displayed to the class and discussed. Questions are typically multiple-choice, although clicker systems which employ students' personal electronic devices allow for short-answer responses as well. Proponents (e.g., Wood, 2004; Premkumar & Coupal, 2008) of clicker systems argue that they have the following advantages:

- The anonymity of clicker responses allows students to respond without fear of embarrassment.
- All students, rather than just a few of the most extroverted and confident, have an opportunity respond to practice problems, making them especially effective in large classes.
- Students who do not understand the topic may realize that others also do not understand.
- Clicker questions serve as mini assessments which improve learning via the testing effect, or the principle that trying to recall information is more beneficial to learning than merely rereading or rehearing it (Goldstein, 2015, p. 186)
- Instructors have immediate feedback regarding student comprehension, which can guide the course of the lesson in real time.
- In general, clicker questions encourage engagement and active learning.

Despite substantial research on implementing clickers in undergraduate-level classrooms, methodological considerations make it difficult to draw conclusions about their efficacy. Researchers often compare classroom approaches which vary not only in whether clickers are used, but also along other dimensions as well. They may compare a lecture approach featuring no practice problems to an approach which features practice problems which students discuss in groups before submitting an answer via clicker. Given the previous discussion of problem-based learning and the forthcoming discussion of group work, it would be no surprise if the second approach outperforms the first, and it is impossible to determine from this research the effect of clickers alone. For instance, Weiss et al. (2020) compared three educational approaches in a thirteen-year study of 1,551 students in undergraduate introductory chemistry. One approach was a traditional lecture class without clickers. The second approach featured a similar lecture, but added practice problems throughout, which students answered via clicker. In this approach, students were allowed, but not required, to discuss with peers before answering. Finally, the third approach included the same practice problems throughout the lecture, but students discussed them in assigned groups of 4 to 6. These approaches were implemented for 5, 14, and 7 semesters respectively, by a single professor, and grading procedures were kept as consistent as possible throughout. The researchers measured student learning through course grades and found that the students in the assigned-group condition, but not those in the clicker-no groups condition, earned statistically significantly higher grades than the students in the traditional lecture condition. These results suggest that merely introducing clicker questions to a lecture format does not increase student performance. Instead, a benefit was seen when both clicker questions and student collaboration were introduced – but importantly, the research does not identify whether the clicker questions brought about a positive effect over and above the known positive effect of student collaboration. Other studies feature similarly confounded designs (e.g., Poirier & Feldman, 2007; Morling et al., 2008; Prezler et al., 2007). Further research is needed which compares teaching approaches which differ only in whether clickers are used.

Although experimental research on the question is lacking, there is some reason to believe that introducing clickers to certain types of classroom activities can improve learning. Buil et al. (2016) propose that providing useful feedback increases student self-efficacy and perceived value of the activity, which leads students to experience more positive emotions during the activity, thereby increasing motivation. When the researchers surveyed students about their attitudes toward small-group clicker activities in an introductory marketing class, their data were consistent with the theoretical model. If this model is accurate, then where clickers are used to provide clear and useful feedback, they are likely to improve student motivation.

While problem-based learning requires open-ended problems, clickers are restricted to multiple-choice or short-answer questions. In my view, cramming PBL activities into a clicker format is likely to reduce the benefit of the former. Instead, clickers are best used in separate activities once content and concepts are introduced. The questions in these activities can enhance understanding of content by recruiting different levels of thinking, ranging from reinforcement of key facts and definitions to the application of knowledge to new situations. The limitations of clickers may make it difficult to engage the highest levels of thinking, like creation and synthesis. Still, a well-crafted question can prompt further discussion which may engage these skills.

Beyond the crafting of focused questions, instructors must make practical decisions regarding how clickers are implemented. Students may be asked to respond to clicker questions alone, or they may be asked to discuss in small groups before responding. Weiss et al. (2020) provide compelling evidence that students benefit most from clickers when they are assigned to small groups. If students work in groups, instructors can decide whether all students respond to the question, or whether the group must come to a consensus and submit only one answer. To my knowledge, there is no research on the effect of this decision on student outcomes, but some research suggests that asking members of a group to share a single clicker encourages collaborative reasoning, and that students prefer this arrangement (McDonough & Foote, 2015). Depending on the options available, the instructor may also need to choose between a system with dedicated clicker devices and a system which employs students' personal devices. Dedicated clicker devices have the advantage of introducing fewer potential distractions to the classroom, while personal devices have the advantage of cost, convenience, and familiarity.

In conclusion, although much of the empirical research does not address the use of clickers in isolation, some evidence and theoretical work suggests that introducing clickers to classroom activities can improve student outcomes. Instructors should carefully consider how to incorporate clickers into their classroom, but in general, clickers are most effective when students work in assigned small groups and each group responds with only a single clicker. Sample clicker questions designed to enhance student understanding of some key topics in introductory logic can be found in Appendix B.

V. Group Activities

The benefits of incorporating group work into classrooms are well-documented. Meta-analyses demonstrate reliable positive effects of group work on student performance in undergraduate classrooms (e.g., Johnson et al., 2014) and undergraduate mathematics specifically (e.g., Springer et al., 1999). In an archetypical study of group work, George (1994) compared student outcomes in two sections of psychology of education. In one section, all students were assigned a partner for the duration of the quarter. In addition to lecture, class sessions included think-pair-share questions and time for students to drill and review content with their partner. In the control session, students were not assigned to groups, and the same activities took place without group work (i.e., instructors posed questions and called on students without offering time for discussion, students were given time to review content individually). Thus, the classroom experience differed only in whether the activities were undertaken individually or in pairs and did not vary in terms of which activities students completed. Additionally, in the cooperative learning condition, students were graded on the success of their pair, while no such grading incentive existed for the control condition. (This decision was based on the research of Slavin (1983), which will be discussed shortly.) All students took the same quizzes and exams. The author found that students in the cooperative-learning condition scored statistically significantly higher than those in the control condition on quizzes and exams, suggesting a clear benefit of introducing group work to the classroom.

Logic instructors have a variety of choices to make when implementing group work. While the literature demonstrates that implementing group work may improve student performance, the specific parameters of group work vary substantially between studies. Johnson et al. (2014) argue that simply assigning groups and instructing students to work together on a task is usually insufficient to achieve the maximum possible benefit of group work. So, what is the best way to incorporate group work into a logic classroom? I recommended that logic instructors include group discussion in problem-based learning and clicker activities. PBL and clicker activities often include group discussion. In group PBL, students work through problems in small groups, sharing ideas, insights, and questions before the full-class discussion (Allen et al., 2001). As mentioned previously, clicker activities are most effective when students discuss questions in small groups and come to a consensus before submitting an answer on a single clicker. Both sorts of activities are therefore excellent opportunities to include group work in the logic classroom.

Instructors must decide how group work will be structured and whether, and if so, how, it will be graded. In an early literature review on group activities in K-12 education, Slavin (1983) categorized research based on the structure of the task: in *task specialization* activities, each group member is responsible for a different portion of the product, while in *group study* activities, group members do not have distinct roles. Then, research was grouped into three categories based on the incentive structure attached to the group work. One structure is *group reward for group products*, where groups are assessed on the quality of a single product (e.g., a worksheet) produced by the group. Another structure is *group reward for individual learning*, where students are assessed individually but are rewarded based on the performance of all group members on the assessment (e.g., each group member receiving a bonus if the mean quiz score of the group exceeds a preset threshold). The final incentive structure is *individual rewards only*, where students may work together to study or complete practice assignments but are ultimately graded only on their own performance on individual assessments. The studies examined in the review showed that group work improves student achievement if and only if its structure is either group study with group reward for individual learning, or task specialization with any group reward structure.

The author concludes from these results that group work improves student achievement if and only if the activity features group rewards and individual accountability. He argues that this is because group rewards and individual accountability create highly motivated groups with norms of success. Task specialization activities promote individual accountability since each member's contribution is both visible to other members and necessary for group success. Group reward for individual learning promotes individual accountability since each group member depends on the achievement of the other members, so they are motivated to identify and help struggling members.

It is unclear how these findings regarding incentive structure can be applied to the sorts of group work I discussed previously. For one, the group-reward-for-individual-learning incentive structure requires consistent, long-term groups. This structure creates accountability by pinning each group member's grade on the success of the other members. However, given the high failure rates seen in introductory logic classes, it would be unreasonable to punish students for being unable to singlehandedly bring their struggling group members back up to speed. Additionally, high drop rates and poor attendance make consistent long-term groups infeasible. Thus, group-reward-for-individual-learning is not an appropriate incentive structure for the types of group activity I have suggested. The benefit of problem-based learning comes from open-ended exploration and discussion, so it does not produce a product which can be evaluated, so group-reward-for-group-product is also inapplicable. In fact, the research on PBL suggests that groups can be productive without any incentive structure (Allen et al., 2001). When it comes to clicker activities with group discussion, clicker responses could be recorded and graded, but if clicker responses are shared with the class in real time, the desire to be correct may be motivation enough. Some task specialization could be created in these activities by assigning one group member to lead the group discussion and another to report to the class discussion (in PBL activities) or enter the group's consensus on a clicker (in clicker activities). In general, it is my view that university students, perhaps in contrast with K-12 students, are mature enough to self-motivate and self-organize during group activities.

One challenge that instructors might face when implementing group work is negative student attitudes. Some students dislike group work, especially if they have previously had negative experiences with it (Wolfe, 1993). Wolfe offers some best practices for implementing group work. Instructors should clearly explain the goals and benefits of group work; structure the group work to promote individual accountability and reduce opportunities for freeloading, which is a major source of frustration for other group members; and schedule all group work during the class period to eliminate the hassle of communicating and scheduling outside of class.

VI. Spaced Practice

In those subjects in which students are taught to solve various types of problems, such as logic, math, physics, and chemistry, many textbooks are organized in roughly the following manner: a lesson on how to solve one type of problem, then a set of problems of that type, then a lesson on how to solve a different type of problem, then a set of problems of that type, and so on for each problem type. Thus, if the students work through the textbook in order, they will practice each type of problem once, in a large block. This practice structure is known as *massed practice*. *Spaced practice* is an alternative to massed practice where students revisit each type of problem at multiple points throughout the term.

When identifying which research on spaced practice was relevant to undergraduate formal logic, two criteria were important. First, much of the research on spaced practice uses inter-practice intervals of less than a day. Such research is illustrative of the benefit of spaced practice in general. However, given that the time between practice sessions (i.e. class meeting and homework assignments) in a college course is typically at least a day, and often multiple days, research which involves inter-practice intervals of around these lengths is most relevant to this investigation. Second, research on spaced practice varies in the type of content being learned. Many studies require participants to memorize and recall verbal information, e.g. nonsense syllables (Ebbinghaus, 1885/1964) or foreign language vocabulary (Bahrick et al., 1993). Those studies in which participants practice problem-solving tasks are most relevant here.

Several studies have demonstrated the benefits of spaced practice on mathematics tasks. In an experiment by Rohrer and Taylor (2006), participants learned a procedure for calculating the number of possible permutations of a set of letters. Then, some participants completed a single practice session consisting of ten problems, while others completed two practice sessions consisting of five problems each, one week apart. The participants in the spaced practice condition scored statistically significantly higher than those in the massed practice condition when tested after four weeks, but not when tested after one week. This suggests that the benefit of spacing practice is moderated by retention interval. (The relationship between inter-practice interval and retention interval has been studied extensively, but not with problem-solving tasks, e.g. Cepeda et al., 2008.) Another experiment by Rohrer and Taylor (2007) used the same permutation task. In this experiment, participants were split into three conditions. Those in the massed practice condition solved four practice problems in a single session. Those in the light massed practice condition solved two practice problems in a single session. Those in the spaced practice condition solved four practice problems, split across two sessions one week apart. When participants were tested one week after their final practice session, those in the spaced practice condition scored statistically significantly higher those in both massed conditions, again suggesting that spacing practice improves learning and retention. Additionally, the test scores of those in the massed and light massed condition were not statistically significantly different. In other words, among those participants who only completed one practice session, doubling the number of problems within that session from two to four did not affect test performance. The authors observe that, while it is possible that increasing the number of practice problems by a greater factor would improve test performance, these results suggest that adding more practice sessions is a more efficient way to improve test performance than increasing the volume of practice in one session.

In the laboratory setting of these studies, practice was spaced by extending the experiment for some participants. (Rohrer and Taylor controlled for this in both of their experiments by requiring massed practice participants to attend dummy sessions.) However, given the amount of content that must be covered in a logic course and the temporal constraints of the academic schedule, it would be impractical to introduce spaced practice to a logic course by only extending the amount of time spent on each type of problem. Instead, introducing spaced practice to a logic course requires that the practice of different problem types overlap. Intuitively, this could divide attention between the various problem types and therefore impede learning. In fact, research suggests that including multiple problem types in the same practice session improves learning. In a second experiment reported by Rohrer and Taylor (2007), participants learned the procedures for calculating the volumes of four geometric solids. Then, they completed two practice sessions, each consisting of sixteen practice problems (four for each geometric solid). For participants in the blocked practice condition, the practice problems were sorted by solid. For participants in the mixed practice condition, the order of practice problems was shuffled. When tested one week later, those in the mixed practice condition scored statistically significantly higher than those in the blocked practice condition. This suggests that mixing problem types within a single practice session can improve learning and retention. However, it may be difficult to leverage this effect in a logic course, since mixing problem types within a single homework assignment may be inconvenient or impractical, especially if the practice problems come from a textbook in which they are already sorted.

At the beginning of this section, I characterized logic textbooks as consisting of lessons and instructions for various problem types, followed by practice problems. My discussion of spaced practice has presupposed that problem types are interchangeable and that, beyond the effect of mixing problem types, the learning and practice of different problem types are independent. Of course, this presupposition is false. Within a logic course, problem types introduced earlier provide the conceptual foundation for those introduced later. For instance, students may first learn to symbolize English arguments in the logical language, then learn to complete proofs, then are asked to symbolize and prove English arguments. It may be unnecessary to revisit problems which only involve symbolization, since the third sort of problems exercises both skills and thus can be considered spaced practice for both. However, some skills and content are not typically naturally spaced throughout the course in this manner. For instance, in *The Power of Logic*, students learn to evaluate arguments for validity, first with truth tables, then with proofs (Howard-Snyder et al., 2020). Of the 33 problem sets in the chapter on proofs, only three require students to complete truth tables. Thus, after the chapter on truth tables, students barely, if ever, revisit the skill. Interspersing practice of truth tables throughout the unit on proofs would improve retention of truth tables. It could also improve learning of proofs by potentially offering insights into how and why the logical system works.

Sometimes, a problem type may be introduced *only* to provide conceptual foundation and practice. For instance, in *The Power of Logic*, students first solve simple proofs with a limited set of inference rules. As more inference rules are introduced, the proofs grow in complexity (Howard-Snyder et al., 2020). Is there a benefit to revisiting those problem types (e.g., simpler proofs with a limited rule set) which are introduced only as stepping-stones to more complex problems? The research on spaced practice does not address this question directly, but the following considerations may be relevant: First, by definition, the desired course outcomes do not include the ability to solve these problems and students are not tested on them, so spending time practicing them may be an inefficient use of time. Also, due to the conceptual relationship between stepping-stone problems and the more advanced problems that follow, it is possible that all skills which would be practiced by revisiting the former are adequately addressed by practicing the latter. However, returning to earlier restrictions could serve as a beneficial challenge which stimulates the student's creativity and flexibility.

The research presented in this section suggests that practicing a skill in multiple spaced sessions rather than a single extended session improves test performance, even as the total number of practice tasks remains constant. While the practice of some skills is naturally interspersed throughout the course, logic instructors should include additional spaced practice of those skills which aren't revisited. Mixing problem types within a single session, where feasible, further improves test performance. Appendix C contains sample syllabi with and without spaced practice, based on the organization of content in *The Power of Logic*.

VII. Conclusion

As in some introductory math courses, a high percentage of students tend to fail introductory formal logic. A common intuition is that reducing class size should improve outcomes in these courses. However, research suggests that merely reducing class size without changing pedagogical approach does not reliably help students. In fact, there are practical changes that instructors can make to the structure of their classrooms to improve student outcomes. Content can be introduced via open-ended problems, encouraging active engagement and metacognition. Clicker questions can then be used to develop student understanding and provide instant feedback to the instructor. Small groups can enhance both types of activity. Finally, instructors can structure their homework assignments to incorporate spaced practice, which keeps all skills and content fresh in students' minds. This combination of modest, evidence-based steps is one small way to improve student outcomes in introductory formal logic.

Appendix A

Sample Problems for Problem-Based Learning

At the beginning of the course, the instructor can introduce foundational concepts by presenting the following two arguments and asking students to evaluate them:

Argument 1:

1. *I (the instructor) am a human being.*
2. *All human beings are mammals.*
3. *So, I am a reptile.*

Argument 2:

1. *I am a honeybee.*
2. *All honeybees are insects.*
3. *So, I am an insect.*

Students may recognize that, although neither argument is convincing, they are unconvincing for different reasons. Specifically, the first argument is unconvincing because, owing to its structure, it is possible for the premises to be true while the conclusion is false. On the other hand, it is not possible for the premises of the second argument to be true while its conclusion is false. The second argument is unconvincing because it has a false premise. This insight lays the groundwork for the concepts of argument form, validity, soundness, and the form/content distinction.

After students have learned to evaluate arguments using truth tables, the instructor can demonstrate the value and conceptual foundation of abbreviated truth tables by presenting the following argument to be evaluated with a truth table, perhaps with a hint to look for a shortcut:

1. $(\sim A \lor D) \lor \sim(\sim B \cdot C)$
2. C
3. $B \rightarrow E$
4. $E \lor \sim D$
5. $A \qquad\qquad \therefore E$

Students may notice that, since two premises are atomic statements, there is no need to include rows in which those atomic statements are false. They may also notice that, since the conclusion is an atomic statement, there is no need to fill out rows in which that atomic statement is true. They may then discover that they can complete what would be a 32-row truth table in only 4 rows.

After students have learned to complete proofs using the implicational rules, the instructor can introduce the rule commutation, and equivalence rules generally, by presenting the following proof:

1. $\sim P \rightarrow (R \vee S)$
2. $T \rightarrow U$
3. $P \vee T$
4. $\sim P \cdot Q$ $\quad\quad\quad\quad \therefore (S \vee R) \cdot U$

The closest that students can get to the conclusion without access to any equivalence rules is: $(R \vee S) \cdot U$. They may realize that commutation is both intuitively valid and required for completing the proof.

Appendix B

Sample Clicker Questions

After students learn the concepts of validity and formal validity, the following questions could be asked to develop their understanding of the distinction between the two:

Which of these arguments are valid? Which are formally valid?

Argument 1:

1. *Alice is older than herself.*
2. *So, Alice is very old.*

Argument 2:

1. *Either a Democrat or a Republican will win the next election.*
2. *No Republican will win the next election.*
3. *So, a Democrat will win the next election*

Argument 3:

1. *Some geometric shapes have four sides.*
2. *So, all squares have four sides.*

Answer: All three arguments are valid. Only argument 2 is formally valid.

After students learn about logically significant categories, the following question could be asked to challenge students to combine their knowledge of those categories in a novel way:

*If **p** is a contingent statement and **q** is a contradiction, then **p** and **q** are:*

A. *Logically consistent and not logically contradictory*
B. *Logically consistent and logically contradictory*
C. *Logically inconsistent and not logically contradictory*
D. *Logically inconsistent and logically contradictory*

Answer: C.

While students are learning to complete proofs, the following question could be asked to clarify some common misunderstandings of inference rules:

Which line of this proof contains an error?

1.	$\sim(A \cdot B) \rightarrow \sim C$	
2.	$\sim(\sim C \vee D)$	$\therefore E \vee A$
3.	$\sim\sim C \cdot \sim D$	2, De Morgan's Laws

4. $C \cdot \sim D$	*3, Double Negation*
5. C	*4, Simplification*
6. $C \rightarrow (A \cdot B)$	*1, Contraposition*
7. $C \rightarrow A$	*6, Simplification*
8. A	*5,7, Modus Ponens*
9. $E \vee A$	*8, Addition*

Answer: Line 7

Line 7 features the use of simplification, an implicational rule, on part of a line. Thus, this question demonstrates an important difference between implicational rules and equivalence rules. Students may incorrectly believe that line 6 or 9 (or some other line) contains the error instead.

Appendix C

Sample Syllabus with Spaced Practice

The following is a comparison of two sample syllabi based on the practice problems and organization of content found in the sixth edition of *The Power of Logic* (Howard-Snyder et al., 2020). The syllabi divide the course into 15 modules, with one homework assignment corresponding to each module. Syllabus A features no explicit spaced practice of problem types, while Syllabus B introduces a moderate degree of explicit spaced practice. Since most problem types build on previous types in some way, I judged that some skills are adequately practiced in Syllabus A and do not need to be revisited in additional practice sessions in Syllabus B. For instance, in Syllabus A, symbolizing is introduced in module 6, and is practiced in modules 8 through 15. Thus, Syllabus B does not include further revisitation of symbolization. Some skills, such as constructing arguments for a given conclusion, are not central to the course, so practice of these skills is not prioritized in Syllabus B. Those skills which are both central to the course and not adequately revisited are marked with an asterisk in Syllabus A. Practice of those skills is spaced among the later homework assignments in Syllabus B. Homework assignments in Syllabus B contain more problem types than those in Syllabus A, but on average, they contain fewer problems, since the largest massed practice sessions in Syllabus A were shortened considerably.

Module	Topics	Syllabus A	Syllabus B
1	Arguments, validity, soundness	1.1C (Informally assessing validity*): 15 problems 1.1D (Informally assessing soundness*): 15 problems	1.1C: 3 problems 1.1D: 3 problems
2	Formal validity, famous forms	1.2B-D (Identifying famous forms*): 30	1.2B-D: 10 1.1C: 3 1.1D: 3
3	Counterexamples	1.3A-D (Constructing counterexamples*): 20	1.3A-D: 5 1.1C: 3 1.1D: 3 1.2B-D: 5
4	Arguments vs. nonarguments	2.1A (Identifying arguments): 20 2.1B (Constructing arguments given a conclusion): 10	2.1A: 20 2.1B: 10 1.2B-D: 5 1.3A-D: 5
5	Well-crafted arguments	2.2A, C (Writing well-crafted arguments*, identifying arguments): 5 2.2B (Writing well-crafted arguments, identifying missing premises): 5 2.2D (Writing well-crafted arguments, identifying famous forms): 5	2.2A, C: 5 2.2B: 5 2.2D: 2 1.1C: 2 1.1D: 2 1.3A-D: 5
6	Symbolizing arguments, logical connectives	7.1A (Identifying well-formed formulas): 10 7.1C-G (Symbolizing): 50	7.1A: 10 7.1C-G: 20 2.2D: 2

Module	Topics	Syllabus A	Syllabus B
7	Truth functions, truth tables	7.2A-B, D (Calculating truth values of compound statements): 20 7.2C (Calculating truth values of atomic statements*): 10	7.2A-B, D: 10 7.2C: 5 2.2D: 2 1.3A-D: 5
8	Evaluating arguments with truth tables	7.3A-B (Truth tables to assess validity*): 15 7.3C (Symbolizing, truth tables to assess validity): 5	7.3A-B: 5 7.3C: 5 1.1C: 2 1.1D: 2 7.2C: 5
9	Abbreviated truth tables	7.4A-D (Abbreviated truth tables to assess validity*): 15 7.4E-F (Symbolizing, abbreviated truth tables to assess validity): 5	7.4A-D: 5 7.4E-F: 5 2.2D: 2 7.3A-B: 3
10	Logically significant categories	7.5A (Truth tables to categorize statements*): 10 7.5B-C (Truth tables to categorize pairs of statements*): 10 7.5D (Symbolizing, truth tables to assess validity): 10	7.5A: 5 7.5B-C: 5 7.5D: 5 1.3A-D: 5 7.4E-F: 3
11	Proofs with eight inference rules	8.1C-E (Direct proof*): 20 8.1F-G (Symbolizing, direct proof): 5	8.1C-E: 15 8.1F-G: 5 7.4A-D: 3 7.5A: 3 7.5B-C: 3

Module	Topics	Syllabus A	Syllabus B
12	Five equivalence rules	8.2C-E (Direct proof): 20 8.2F (Symbolizing, direct proof): 5	8.2C-E: 15 8.2F: 5 1.1C: 2 7.3A-B: 3 7.5A: 3 7.5B-C: 3
13	Five more equivalence rules	8.3C-E (Direct proof): 20 8.3F (Symbolizing, direct proof): 5	8.3C-E: 15 8.3F: 5 1.1D: 2 2.2D: 2 7.4A-D: 3
14	Conditional proof (CP)	8.4A-B (Proof with CP): 10 8.4C-D (Symbolizing, proof with CP): 5 8.4E (Symbolizing, abbreviated truth tables to assess validity, proof with CP): 5	8.4A-B: 10 8.4C-D: 5 8.4E: 5 1.3A-D: 5 8.3C-E: 5 7.5A: 3
15	Reductio ad absurdum (RAA)	8.5A, C (Proof with RAA): 10 8.5B (Abbreviated truth tables to assess validity, proof with RAA and CP): 5 8.5D-E (Symbolizing, proof with RAA): 5 8.5F (Symbolizing, abbreviated truth tables to assess validity, proof with RAA and CP): 5	8.5A, C: 10 8.5B: 5 8.5F: 5 7.3A-B: 3 8.3C-E: 5 7.5B-C: 3

References

Allen, D.E., Duch, B.J., & Groh, S.E. (2001). Strategies for using groups. In B.J. Duch, S.E. Groh, & D.E. Allen (Eds.), *The power of problem-based learning: A practical "how to" for teaching undergraduate courses in any discipline* (pp. 59-68). Stylus Publishing.

Bahrick, H.P., Bahrick, L.E., Bahrick, A.S., & Bahrick, P.E. (1993). Maintenance of foreign-language vocabulary and the spacing effect. *Psychological Science, 4*, 316-321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

Bazzaz, D. (2022, January 18). Washington students' test scores drop significantly in first exams since pandemic began. *The Seattle Times*. https://www.seattletimes.com/education-lab/in-first-assessment-since-the-pandemic-began-washington-student-test-scores-drop-significantly/

Biddle, B.J., & Berliner, D.C. (2002). Small class size and its effects. *Educational Leadership, 59*(5), 12-23. https://www.ascd.org/el/articles/small-class-size-and-its-effects

Buil, I., Catalán, S., & Martínez, E. (2016): Do clickers enhance learning? A control-value theory approach. *Computers and Education, 103*, 170-182. https://doi.org/10.1016/j.compedu.2016.10.009

Cepeda, N.J., Vul, E., Rohrer, D., Wixted, J.T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095-1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development, 21*(1), 5-21. https://www.ingentaconnect.com/contentone/magna/jfd/2007/00000021/00000001/art00001

Dolmans, D.H.J.M., De Grave, W., Wolfhagen, I.H.A.P., & van der Vleuten, C.P.M. (2005). Problem-based learning: Future challenges for educational practice and research. *Medical Education, 39*(7), 732-741. https://doi.org/10.1111/j.1365-2929.2005.02205.x

Duch, B.J. (2001). Models for problem-based instruction in undergraduate courses. In B.J. Duch, S.E. Groh, & D.E. Allen (Eds.), *The power of problem-based learning: A practical "how to" for teaching undergraduate courses in any discipline* (pp. 39-45). Stylus Publishing.

Duch, B.J., Groh, S.E., & Allen, D.E. (2001). Why problem-based learning? A case study of intsitutional change in undergraudate education. In B.J. Duch, S.E. Groh, & D.E. Allen (Eds.), *The power of problem-based learning: A practical "how to" for teaching undergraduate courses in any discipline* (pp. 3-11). Stylus Publishing.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H.A. Ruger & C.E. Bussenius, Trans.). Dover. (Original work published 1885)

Fletcher, D. (2018). *Global problem-based learning in math* (Publication no. 4939) [Doctoral dissertation, University of South Carolina]. University of South Carolina Scholar Commons.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., & Wenderoth, M.P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

George, P.G. (1994). The effectiveness of cooperative learning strategies in multicultural university classrooms. *Journal on Excellence in College Teaching, 5*(1), 21-30. http://celt.miamioh.edu/ject/issue.php?v=5&n=1

Goldstein, E.B. (2015). *Cognitive psychology: Connecting mind, research, and everyday experience* (4th ed.). Cengage Learning.

Howard-Snyder, D.T., Howard-Snyder, F.M., & Wasserman, R.J. (2020). *The power of logic*. (6th ed.). McGraw Hill.

Jarvis, T.J. (2007). Class size and teacher effects on student achievement and dropout rates in university-level calculus. Unpublished manuscript. Retrieved from http://www.math.byu.edu/~jarvis/

Johnson, D.W., Johnson, R.T., & Smith, K.A. (2014). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching, 25*(3-4), 85-118. http://celt.miamioh.edu/ject/issue.php?v=25&n=3%20and%204

McDonough, K. & Foote, J.A. (2015). The impact of individual and shared clicker use on students' collaborative learning. *Computers & Education, 86*, 236-249. https://doi.org/10.1016/j.compedu.2015.08.009

McMurtrie, B. (2022). A 'stunning' level of disconnection: Professors are reporting record numbers of students checked out, stressed out, and unsure of their future. *The Chronicle of Higher Education, 68*(17), 12. https://www.chronicle.com/article/a-stunning-level-of-student-disconnection

Morling, B., McAuliffe, M., Cohen, L., & DiLorenzo, T.M. (2008). Efficacy of personal response systems ("clickers") in large, introductory psychology classes. *Teaching of Psychology, 35*(1), 45-50. https://doi.org/10.1080/00986280701818516

Olson, J.C., Cooper, S., & Lougheed, T. (2011). Influences of teaching approaches and class size on undergraduate mathematical learning. *PRIMUS, 21*(8), 732-751. https://doi.org/10.1080/10511971003699694

Pearlman, S. (2019). Flipping the logic classroom. *Teaching Philosophy, 42*(4), 355-373. https://doi.org/10.5840/teachphil20191018112

Poirier, C.R., & Feldman, R.S. (2007). Promoting active learning using individual response technology in large introductory psychology classes. *Teaching of Psychology, 34*(3), 194-196. https://doi.org/10.1080/00986280701498665

Premkumar, K., & Coupal, C. (2008). Rules of engagement: 12 tips for successful use of "clickers" in the classroom. *Medical Teacher, 30*(2), 146-149. https://doi.org/10.1080/01421590801965111

Prezler, R.W., Dawe, A., Shuster, C.B., & Shuster, M. (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE: Life Sciences Education, 6*(1), 29-41. https://doi.org/10.1187/cbe.06-09-0190

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*(9), 1209-1224. https://doi.org/10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*(6), 481-498. https://doi.org/10.1007/s11251-007-9015-8

Shipman, H.L., & Duch, B.J. (2001). Problem-base learning in large and very large classes. In B.J. Duch, S.E. Groh, & D.E. Allen (Eds.), *The power of problem-based learning: A practical "how to" for teaching undergraduate courses in any discipline* (pp. 149-163). Stylus Publishing.

Slavin, R.E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94*(3), 429-445. https://doi.org/10.1037/0033-2909.94.3.429

Springer, L., Stanne, M.E., & Sonovan, S.S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21-51. https://doi.org/10.2307/1170643

Strobel, J., & van Barneveld, A. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-Based Learning, 3*(1), 44-58. https://doi.org/10.7771/1541-5015.1046

Weiss, J.T., McGuire, P., Clouse, W., & Sandoval, R. (2020). Clickers are not enough: Results of a decade-long study investigating instructional strategies in chemistry. *Journal of College Science Teaching, 49*(3), 58-65. https://www.jstor.org/stable/27045865

When all else fails: Student failure rates in required courses continue to skyrocket. (2017, July 18). *The Western Front.* https://www.westernfrontonline.com/article/2017/07/when-all-else-fails-student-failure-rates-in-required-courses-continue-to-skyrocket

Wolfe, A. (2012). Implementing collaborative learning methods in the political science classroom. *Journal of Political Science Education, 8*(4), 420-432. https://doi.org/10.1080/15512169.2012.729451

Wood, W.B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell, 7*(6), 796-798. https://doi.org/10.1016/j.devcel.2004.11.004

Yew, E.H.J., & Goh, K. (2016). Problem-based learning: An overview of its processes and impact on learning. *Health Professions Education, 2*(2), 75-79. https://doi.org/10.1016/j.hpe.2016.01.004