Methods in Physiology
Published on behalf of The American Physiological Society by Springer

Harry J. Witchel
Michael W. Lee *Editors*

Technologies in Biomedical and Life Sciences Education

Approaches and Evidence of Efficacy for Learning

american physiological society

Springer

Methods in Physiology

Published on behalf of The American Physiological Society by Springer

This series describes experimental techniques in cellular, molecular and general physiology. Each book is edited by experts in their field and covers theory and history behind the method, critical commentary, major applications with examples, limitations and extensions of each technique, and vital future directions.

Harry J. Witchel • Michael W. Lee
Editors

Technologies in Biomedical and Life Sciences Education

Approaches and Evidence of Efficacy for Learning

Springer

american
physiological
society

Editors
Harry J. Witchel
Department of Neuroscience
Brighton and Sussex Medical School
Brighton, UK

Michael W. Lee
Department of Medical Education
Geisel School of Medicine, Dartmouth College
Hanover, NH, USA

ISSN 2628-7471 ISSN 2628-748X (electronic)
Methods in Physiology
ISBN 978-3-030-95632-5 ISBN 978-3-030-95633-2 (eBook)
https://doi.org/10.1007/978-3-030-95633-2

Published on behalf of the American Physiological Society.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Part II
How Educational Technologies Shape the Classroom Experience

Chapter 5
Perceptual Learning, Adaptive Learning, and Gamification: Educational Technologies for Pattern Recognition, Problem Solving, and Knowledge Retention in Medical Learning

Philip J. Kellman, Victoria Jacoby, Christine Massey, and Sally Krasne

Abstract In this chapter, we consider recent advances in the learning sciences and their potential and actual applications in medical learning. We describe emerging ideas that broaden traditional declarative and procedural emphases in learning, with a special focus on *perceptual learning*—experience-induced improvements in the pickup of information. Experience in a task or domain changes perception to be more efficient at discovering and extracting relevant information and ignoring extraneous information while also reducing effort and cognitive load. These discovery and fluency changes from perceptual learning are crucial contributors to complex task performance and advanced expertise. We also examine adaptive learning and its connections to scientific research on the testing effect and the spacing effect, as well as other examples of technology-enhanced learning (TEL), including games and gamification in learning generally and medical education specifically. New learning

Funding Announcement The authors gratefully acknowledge research support from NIH/NCI grant 5R01CA236791 to PJK, SK, and CM; NSF ECR award 1644916 to PK and CM; and support from the US Office of Naval Research Cognitive Science of Learning, US Military JPC-1/MSIS, and US Army RDECOM programs.

P. J. Kellman (✉)
Department of Psychology, University of California, Los Angeles, Los Angeles, CA, USA

Department of Surgery, UCLA David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
e-mail: kellman@cognet.ucla.edu

V. Jacoby · C. Massey
Department of Psychology, University of California, Los Angeles, Los Angeles, CA, USA
e-mail: vjacoby@g.ucla.edu; cmassey@psych.ucla.edu

S. Krasne
Department of Physiology, UCLA David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
e-mail: skrasne@g.ucla.edu

H. J. Witchel, M. W. Lee (eds.), *Technologies in Biomedical and Life Sciences Education*, Methods in Physiology, https://doi.org/10.1007/978-3-030-95633-2_5

technologies combining perceptual and adaptive learning make it possible to accelerate perceptual learning and rapidly advance aspects of expertise, such as pattern recognition, that have been elusive in most instructional contexts. We illustrate the efficacy of research-informed approaches to TEL in our work applying perceptual and adaptive learning modules (PALMs) in medical learning domains. We outline the scope of this research and illustrate characteristic elements and findings of PALMs using the example of electrocardiography. We also describe research that suggests that perceptual and adaptive learning in PALMs and related efforts have great potential to improve medical learning, not only for difficult perceptual classifications, but for factual learning and higher-order diagnostic skills as well.

5.1 Introduction

It is an exciting time to be involved in medical learning, either as a researcher or a practitioner seeking to improve it. Efforts in cognitive psychology and related disciplines to understand and improve learning—broadly described as the "science of learning"—have expanded dramatically in recent years. These efforts have yielded many new findings of practical importance such as spacing and interleaving in learning and the advantages of testing over studying (e.g., Carpenter et al., 2012; Pashler et al., 2007). Some of them have also begun to expand and change our very conceptions about what learning is (Kellman & Massey, 2013), a transformation that offers previously unanticipated opportunities for enhancing instruction with new methods that accelerate the development of expert pattern recognition.

In this chapter, we consider recent advances in the learning sciences and their potential and actual applications in medical learning. We describe concepts that broaden traditional declarative and procedural emphases in learning, with a special emphasis on *perceptual learning*—experience-induced improvements in the pickup (extraction and differentiation) of information (Gibson, 1969; Kellman & Garrigan, 2009). We also focus on adaptive learning and its connections to scientific research on the testing effect, the spacing effect, motivation, gaming, and mastery learning. We review efforts to apply perceptual and adaptive learning, as well as game concepts, in medical education. Following this review, we will focus in some detail on our work on adaptive perceptual learning. We use our experiments training medical students and residents to interpret ECGs to illustrate, in detail, both our approach and the types of findings we have made in a variety of areas, both published (histopathology, dermatology, echocardiography and electrocardiography) and only presented thus far at professional meetings (obstetrics and gynecology, radiology, point-of-care ultrasound, ophthalmology). We close with a discussion of issues and prospects for the further development of advanced learning technologies in medical learning.

5.2 Conceptions of Learning

When we think of what "learning" is, several prototypical examples come to mind. The storage of information in memory is one; learning a sequence of steps in a procedure is another. These declarative and procedural kinds of learning are important, but learning includes much more. One way to appreciate other crucial elements is to look backward from the long-term goals of learning. Consider, for instance, a newly minted MD's ability to interpret an abnormal ECG, detect a suspicious calcification on a mammogram, or pick up a subtle heart murmur using a stethoscope. Even the most diligent of students will nevertheless be slow and prone to making errors—in stark contrast to the seemingly effortless performance of an experienced practitioner. Although we dedicate intensive instructional effort to written materials and classroom time ostensibly to "teach" and develop competency in domains as diverse as medical diagnosis, aviation training, chemistry, and mathematics, it is accepted and obvious that when coursework and examinations end, even the student who has excelled in learning the facts and procedures is not a pilot, chemist, surgeon, or diagnostician. That functional expertise in most domains requires further extended experience, such as multi-year medical residency, speaks to other ingredients needed to fully "learn" or achieve real competence.

5.2.1 Perceptual Learning

What is missing from typical ideas of learning and most learning research? Oddly, if one looks at the different but related domain of research on expertise, one does not see this focus on declarative and procedural knowledge. Studies of expertise consistently identify as the most striking difference between experts and novices, not their storehouse of information, not their procedural knowledge, but their *information extraction and pattern recognition*. Referring to classic research on chess players by De Groot (1965), an influential review of research on learning and education (Bransford et al., 1999) put it this way:

> Experts are not simply "general problem solvers" who have learned a set of strategies that operate across all domains. The fact that experts are more likely than novices to recognize meaningful patterns of information applies in all domains, whether chess, electronics, mathematics, or classroom teaching. In DeGroot's (1965) words, a "given" problem situation is not really a given. Because of their ability to see patterns of meaningful information, experts begin problem solving at "a higher place." (DeGroot, 1965)

These changes derive from a set of learning processes called *perceptual learning*, broadly defined as experience-induced improvements in the extraction of information.

The idea that perceptual pickup of information changes with experience in a task or domain may be unfamiliar to many. That perceptual learning attunes perception to discover and extract meaningful information, often involving complex relations, may be similarly novel. Traditional views of perception as static, elementary, or providing essentially meaningless raw material unless supplemented by other cognitive processes would not suggest these ideas. However, contemporary research and theory in perception embraces a very different view of the outputs of perception. Perception is oriented toward extracting meaningful structure of objects and events; it changes dynamically to optimize task performance; and it interacts closely with other cognitive processes. (For discussion, see Kellman & Massey, 2013.)

Indeed, even the declarative and procedural aspects of learning, which might seem "post-perceptual," are deeply connected to pattern recognition skills developed through perceptual learning. Vast storehouses of declarative information, or learned procedures, are nearly useless without the learner being able to tell which facts, concepts, or procedures apply to the case or situation one is confronting. Classifying a situation, interpreting an image, and recognizing a constellation of symptoms—these are all pattern recognition problems, and the way these pattern recognition abilities develop and improve is through perceptual learning. As an example, one of the authors is a pilot who owned an aircraft with a six-cylinder piston engine. One day, as the pilot taxied in after a flight and parked at his home airport, the chief mechanic happened to be walking by. After the pilot got out of the plane, the mechanic said, "How long have you had that dead cylinder?" The pilot said "What dead cylinder?" The engine sounded normal to the pilot, but the mechanic could hear in a few seconds that one of the six cylinders had failed completely! Obviously, had the pilot noticed a significant power loss in flight and later asked the mechanic about possible causes, the troubleshooting process would have been rather greatly accelerated by the entry point provided by the mechanic's expert ear, as opposed to tearing down the whole engine trying to discover a problem. We have no doubt that medical practitioners in various specialties could recount a broad range of analogous anecdotes.

The neglect of perceptual learning in most accounts and in instructional practice relates to characteristic difficulties in teaching and learning. Perceptual learning grows gradually and often implicitly, through encountering instances and making classifications. Unlike the explicit teaching of a fact or procedure to able students, as in medical learning contexts, this form of learning is not advanced much by lectures or even demonstrations, and much of what is learned is often not verbalizable by instructors or students (e.g., Benzil, 2018). These aspects of perceptual learning account for why this kind of learning languishes under traditional approaches, and they are key to the relevance of long apprenticeship or experience to reach expert levels of pattern recognition.[1] More direct attention to perceptual learning in training has the potential to accelerate this process dramatically.

---

[1] It is important to note that scientific work on perceptual learning is distinctly different from sociocultural and educational theories sometimes referred to as experiential or situated learning (e.g., communities of learners (Rogoff, 1994), situated learning (Lave & Wenger, 1991), and activity theory (Engeström et al., 1999), among others). Although both perceptual learning and sociocultural experiential learning theories emphasize the importance of extended experience to promote learning, the learning processes and outcomes that they are concerned with diverge

We emphasize the pattern recognition aspects of perceptual learning because they are important and somewhat intuitive to understand. In reality, the ways that perception improves with experience are various, and together these perceptual learning effects exert even broader and more profound effects on expertise. Perceptual learning effects fall into two broad categories, which have been called *discovery* and *fluency* effects (Kellman, 2002). *Discovery* effects refer to learners coming to find and select the information that is relevant to a task—for example, the features or patterns that distinguish instances of one category from another. Learners come to extract relevant information, including new relationships or patterns to which they were originally insensitive (Gibson, 1969; Goldstone, 2000). *Fluency effects* refer to improvements in the efficiency of information extraction. Whereas the "what" (or where, etc.) in information pickup is the province of discovery, fluency involves processes that expedite the pickup of relevant information. Experience or practice leads to perceptual extraction that is faster, involves more parallel processing, picks up larger "chunks" of information, and occurs with less effort or cognitive load. This latter characteristic—greater automaticity in information extraction—allows the operator to use attentional resources for other components of complex tasks and situations. When discovery and fluency effects are considered together, it becomes clear why there is a gap between a learner's having stored relevant factual, conceptual, and procedural information and being an expert. In fact, there is more than one gap. The activation of relevant information or actions through pattern recognition is crucial in the initial grasp of a problem but also continues at each step in a procedure, as information acquired guides the choice or nuances of each new action. At the same time, new actions produce new information. This interleaving of information pickup and action has been recognized in the concept of "perception-action loops" (Gibson, 1966). The discovery aspects of perceptual learning lead to more selective pickup of relevant information and sensitivity to new relations at many points in a task. Even in tasks that may not primarily involve perception-action loops, such as medical image interpretation, growing expertise leads to better actions in terms of where to look; what to scrutinize or magnify, etc.; what to think about; and what to do next based on information acquired. In complex situations, fluency effects allow coordination of various task components, including the guiding of procedures with reduced attentional load, allowing more effective management of multiple task demands and communication, as well as better situational awareness. The apparent magic of advanced expertise in many domains (chess, music, science, medicine)

considerably. Sociocultural learning theories are primarily focused on the learning of community norms, values, and identities, as well as workplace roles and practices as they are situated in particular contexts. They emphasize interpersonal interactions and collaborative activities as the means by which learning is structured and mediated (Yardley et al., 2012). In contrast, perceptual learning involves changes in the operation of perceptual systems that lead to domain-specific selection and more rapid pickup of relevant information, extraction of relevant patterns or relations, and reductions in effort or attentional load in perception. It is likely that experiential learning and perceptual learning co-occur. Workplace or apprenticeship learning in medical education presumably provides conditions that can support both of these types of learning.

owes a great deal to the expert's specialized, advanced, and automatic information extraction, continuously interacting with other cognitive and procedural demands of complex tasks (Bryan & Harter, 1899; Kellman & Garrigan, 2009).

Broadening ordinary conceptions of learning, and specifically the understanding perceptual learning, offers clear opportunities to improve medical learning and the acceleration of expertise. This is especially true when training approaches combine perceptual learning with adaptive learning methods, to which we now turn.

5.3 Adaptive Learning

Adaptive learning may be defined broadly as the use of one or more aspects of a learner's performance to arrange learning content and events. Early work often referred to such arrangements as *adaptive instructional systems* (e.g., Atkinson, 1976), but the more or less synonymous phrase adaptive learning has become more common. There can be many forms of adaptivity, ranging from simple pre-testing to determine a learner's starting point, modification of learning content to address each learner's strengths and weaknesses, or trial-by-trial tracking of learning performance to continually decide the sequence of learning events in a learning session. While various forms of teacher-led differentiation of instruction for different learners have been in widespread use for some time (Bondie et al., 2019), possibilities and investigations of adaptive methods have blossomed with the emergence of generally available digital technologies which have the speed and capacity to arrange learning events in real time based on individual learner responses and to store large databases of learning materials.

Recently, medical education specifically has embraced emerging technology-enhanced learning or *TEL*. It has been argued that these efforts could be further enhanced by greater use of adaptive learning (Sharma et al., 2017). A number of efforts incorporating adaptive learning in medical education have been reported, spanning a range of topic areas and types of learning (Kellman & Krasne, 2018; McMahon & Drazen, 2014). One example is the NEJM Knowledge+ "Pain Management and Opioids" learning module (Hamnvik et al., 2019), which tests learners and takes both learner accuracy and rated confidence as inputs to determine gaps in knowledge and trigger presentation of learning content most needed. Hamnvik et al. (2019) cite a conference paper (Healy et al., 2019) reporting that use of the NEJM Knowledge+ review platform lowered the failure rate on the American Board of Internal Medicine (ABIM) maintenance of certification exam from 11 to 5%.

Use of the term "adaptive" or adaptive features is not a guarantee of learning effectiveness when compared to nonadaptive, business-as-usual control groups. Griff and Matter (2013) measured the effectiveness of adaptive learning in anatomy and physiology courses. Multiple schools were recruited for this study, and within each school students were assigned to either an adaptive learning or control condition. In the adaptive learning condition, students answered multiple choice, multi-select, and fill-in-the blank questions on a system called "LearnSmart" developed by

McGraw Hill Higher Education. LearnSmart uses the accuracy of the answers given in combination with each student's confidence rating to select the next questions that the student sees. Ideally, this should allow students to spend more time and receive more practice on topics that are more difficult for them. In the control condition, students were given practice questions chosen for them by their instructor from a larger question bank supplied by the class's textbook publisher. All students were given a pre-test and a post-test, both containing the same questions. The duration between the pre- and post-test varied. The study was conducted across six institutions with the pre-test occurring at the beginning of the quarter or semester, and the post-test was taken at the end of that quarter/semester.

Ultimately, the results showed that despite positive student feedback about the LearnSmart program, in most schools there was no significant difference between the adaptive and control groups in student performance gains from the pre-test to post-test. The result is consistent with a number of studies in the emerging judgments of learning (JOL) field showing that learner judgments of confidence are influenced by numerous factors that may not be predictive of actual learning or future performance (Kornell & Hausmann, 2017).

More generally, as with TEL, adding in adaptive or technological elements does not guarantee success. Spector (2014) emphasizes that there is often an eagerness to adopt the technology without thoroughly assessing its capabilities or without serious consideration for how it is integrated into a classroom. Similar comments apply to the use of games in learning. (See below.) Another issue is that many efforts in adaptive learning have been evaluated through user satisfaction measures rather than through objective measures of learning and retention.

5.3.1 Learning Principles and Adaptive Learning

We think that adaptive learning efforts are most likely to succeed when they are based on learning principles that have been clearly tested and supported in research using objective measures of learning. In this section, we briefly describe two such principles, both as examples of research-based concepts that may be valuable and as ideas that have been implemented in some successful adaptive learning approaches. These are the *testing effect* and the *spacing effect*.

5.3.1.1 The Testing Effect

Testing is usually done to assess learning. Learners typically prepare for testing by studying. It is mildly paradoxical, then, that testing events themselves have been shown to improve learning more so than studying. This idea, which has a long history (e.g., Abbott, 1909; Tulving, 1967), has come to be known as the *testing effect* (Bjork, 1975; Carrier & Pashler, 1992; Roediger & Karpicke, 2006).

The testing effect has been demonstrated in both laboratory and applied contexts (Roediger & Butler, 2011). In one example, Larsen et al. (2013) measured the testing effect in a real-world medical learning setting using residents attending an instructional conference. All residents participated in an hour-long teaching session on two topics: *status epilepticus* and *myasthenia gravis*. Following the session, half of the residents were given a test on *status epilepticus* and a review sheet to study for *myasthenia gravis*; the other half were given a test on *myasthenia gravis* and a review sheet to study for *status epilepticus*. Both the tests and review sheets covered the same information, and residents were able to check their answers following the test. This process was repeated approximately 2 weeks later, and at 6 months, all residents returned to take a final test on both topics. The results showed that despite being given the same information and amount of time devoted to each topic, residents who engaged with the material through tests significantly outperformed those who studied with the review sheet for each topic.

Why testing events are superior to studying events is not entirely settled. One likely factor is that testing events activate more attention and effort. A "retrieval effort hypothesis" suggests that more difficult (but successful) retrievals of information provide greater gains in long-term retention (Pyc & Rawson, 2009). Studying does not exercise retrieval processes in the way that testing does. Testing events with feedback may also help learners to recognize deficiencies in their current understanding of material (Roediger & Butler, 2011), whereas restudying may give illusions of familiarity that lead learners to overestimate what they know about the topic (Bjork & Bjork, 2011). Other contributors to testing effects may be greater elaboration and connections of knowledge that occur during retrieval and the idea that testing practice resembles actual test situations more so than passive studying.

Adaptive learning and the testing effect go together naturally. Adaptivity always depends in some way on assessment of a learner's status or needs. Although some adaptive efforts may use initial testing to choose a level or starting point for learning, testing and adaptive learning can be more richly combined when most of the learning itself consists of ongoing testing events (e.g., Mettler et al., 2016). With the capacities of modern computer systems, such arrangements can acquire rich and up-to-date estimates of learner performance and use the data to guide learning events in productive ways. Prominent among these is to guide spacing in learning.

5.3.2 The Spacing Effect

Considerable research in the learning sciences converges in showing that spacing the repetitions of an item in learning, rather than grouping them in close succession, enhances long-term retention. This benefit generalizes across numerous tasks, materials, and time scales and has emerged as one of the most robust effects in learning and memory research (Bjork & Bjork, 2011; Karpicke & Roediger, 2007). Although most spacing research has focused on declarative (factual) learning materials, as in foreign language vocabulary, paired associates, or learning facts from text, spacing research has been known for some time to have benefits in motor learning (Lee & Genovese, 1988), although this may vary by task (Wiseheart et al., 2017). More recently, spacing has been shown to have powerful effects in perceptual learning (Mettler & Kellman, 2014). Thus, spacing considerations are likely to be important across different types of learning, and they offer fertile opportunities for enhancing learning using instructional technology.

Almost certainly, the explanation for spacing effects involves multiple factors (Glenberg, 1979; Mettler et al., 2016). A fundamental one involves the well-studied differences between massed and distributed practice. Revisiting a learning item several times in succession offers little or no benefit for long-term learning. The problem is that in such massed practice, relevant information remains in working memory, providing no practice in retrieval from more long-term memory (Bjork & Bjork, 2011). An analogous problem occurs in more complex tasks. In perceptual-motor tasks, for example, learners retain components of performance and need not organize the full processing and response selection as would be required with more widely spaced challenges (Schmidt & Bjork, 1992). In perceptual learning, or the aspects of complex tasks that involve perceptual learning, with massing, the learner may remain attuned from one trial to the next to particular information useful for classification.

Requiring that each learning event "stands on its own," free of transient carryover from recent events, is not the only driver of spacing benefits. Spacing of items usually coincides with interleaving, such that between repetitions, not only does time elapse but other learning items occur. Interleaving exerts benefits by connecting retrieval to changing contexts and by facilitating discrimination between varying items that occur nearby in time (Kang & Pashler, 2012; Kornell & Bjork, 2008).

A considerable amount of research has attempted to establish principles for predetermined spacing schedules in learning. Some evidence suggests that expanding recurrence intervals outperform fixed ones (e.g., Landauer & Bjork, 1978). Other work suggests that equal interval practice may actually be superior to expanding practice intervals when measured after a delay (Karpicke & Roediger, 2007). On the other hand, Storm et al. (2010) reported better memory from an expanding schedule, but only when intervening items during spacing were highly related to spaced items, suggesting that expanding intervals work best for difficult material.

Mettler et al. (2016) argued that the question of the superiority of fixed vs. expanding spacing has no ultimate answer. There are many factors known to relate to optimal spacing. Most, if not all, exert their effects on an internal variable of learning strength. In general, the optimal time to practice an item is when retrieval is difficult but can still succeed. This "retrieval effort hypothesis" (Pyc & Rawson, 2009) fits a great deal of research (Bjork & Bjork, 1992; Johnston & Uhl, 1976; Pyc & Rawson, 2009; Thios & D'Agostino, 1976). Retrieval effort decreases as learning strength increases. The problem in schemes that preset spacing intervals is that fluctuations of learning strength in a learning session arise from numerous and subtle influences that cannot be captured in an a priori model. Certainly, learning strength tends to decrease over time without practice, and practice events (especially testing events) tend to increase it; but in any set of learning items, difficulty will vary by item, by learner, and by interactions of the two. Many events that occur during the course of a learning session, such as attentional fluctuations, momentary confusions between items, or resolution of confusions by patterns of interleaving and feedback, may affect a given learner's learning strength for a given item.

5.3.3 Adaptive Learning and the ARTS System

These issues can be addressed by adaptive methods that assess learning strength in an ongoing manner and adjust spacing intervals accordingly. Adaptive learning systems have generally relied on accuracy data alone to assess progress, but accuracy information alone is too crude to serve as the basis of optimal spacing. Several years ago, we proposed that response times, along with accuracy, can better assess underlying learning strength and serve as the basis of spacing. We developed the ARTS (Adaptive Response-Time-based Scheduling) system to do this (Kellman, 2006; Kellman & Krasne, 2018; Mettler et al., 2011, 2016). Actually ARTS adds response times to adaptive systems in two different ways: One is spacing, but the other innovation is using response time in combination with accuracy to set mastery criteria, an especially useful arrangement for many learning domains where both accuracy and fluency are goals of learning.

The ARTS system has proven to be highly effective in various domains, especially in STEM learning and medical learning, and has been shown to outperform a classic adaptive learning system (Mettler et al., 2011). These effects have been evident in both factual and perceptual learning domains, suggesting that a *successful effort hypothesis* may hold across various types of learning, generalizing the retrieval effort hypothesis to apply to types of learning that do not depend primarily on retrieval of memory items. Below we consider especially the application of ARTS to perceptual learning in perceptual-adaptive learning modules applied to a variety of medical learning domains.

With regard to efforts to optimize spacing, something more than evidence of effectiveness is required. Mettler et al. (2016) performed experiments whose results confirm the idea that optimal spacing intervals must be adaptive and based on the individual learner. They first showed the superiority of adaptive learning to a nonadaptive control condition. They then compared adaptive learning conditions to two "yoked" control conditions. In a *yoked-random* condition, a new group of participants received the same spacing intervals that had been generated by adaptive participants, but randomized across the learning items. If the benefit of adaptive learning derives, not from individually adaptive intervals per se, but from the overall distribution of spacing intervals that occur under adaptive conditions, then this group would have been expected to perform equally well as the adaptive condition. In a *yoked-item* condition, spacing schedules that had been generated by adaptive participants remained attached to the same learning items. If the benefit of adaptive conditions derives from some items consistently having greater difficulty than others, then this group would have been expected to match the performance of adaptive participants.

The adaptive learning condition outperformed both yoked control conditions. There was a trend for the yoked-item condition to be better than the yoked-random condition, suggesting that item difficulty does have an influence and that such consistent effects of difficult and easy items are appropriately picked up by ARTS and used efficaciously in spacing. Ultimately, however, adaptive intervals generated by each learner's own performance give the best learning results. These results are consistent with the idea that optimal spacing must be tied to individual learner's learning strength for learning items, and they also confirm the importance of the underlying construct of learning strength as the driver of optimal spacing. Predetermined (nonadaptive) spacing arrangements cannot achieve the same results.

5.4 Games and Gamification in Medical Learning

Earlier we noted the role of digital technology as empowering the development and implementation of adaptive learning methods. Similarly, the widespread availability of computers has opened up new opportunities for educational games and gamification. *Gamification* refers to the application of game elements to non-game contexts (Deterding et al. 2011; Hamari et al., 2014; Rutledge et al., 2018).

The motivations for making learning games, or making learning activities more like games, are several. In domains where many students struggle, such as mathematics and other STEM subjects, it has been hoped that games or gamification might increase student attention, engagement, and motivation. Instructors note with envy the degree of immersion and engagement that is evident when students play video games. If there were a way to engender such levels of attachment and commitment in academic subjects, it might boost learning outcomes. If the effect of gamification were simply to increase the time learners spend on learning tasks, it might plausibly advance learning.

An interesting preliminary consideration is that it is hard to define precisely what is meant by a game, or even game-like elements. Strikingly, the philosopher Wittgenstein used “game” as a prime example in making the case that human concepts do not have precise definitions or meanings (Wittgenstein, 1953). That said, some game-like features that have been considered to be potentially useful in learning include competition, speeded responses, feedback, markers of progress, ascent of levels with increasing success, and well-defined rewards. Obviously, some of these elements have been used in learning settings apart from explicit efforts to make game-like activities.

The idea that games might help in serious learning contexts has had remarkable momentum, especially when contrasted with the evidence base. The widespread enthusiasm for the use of learning games or gamification has been primarily driven by intuition (and, perhaps, advertising). Actual research on use of games has indicated few clearly positive results when assessed by objective measures of learning. For example, a broad review by Young et al. (2012) found no compelling evidence for learning benefits of games in STEM domains but possibly a benefit for a small number of games in social science domains. A more recent meta-analysis (Sailer & Homner, 2020) found modest positive evidence for gamification effects on learning but reported inconsistencies in which elements lead to successful outcomes.

In medical learning applications, gamification has similarly led to mixed results and conflicting views. One concern is that with many potential game elements to choose from, an overall decision to "gamify" is less meaningful than specification of particular features. Research on these specific factors seems more likely to be meaningful and transferable than the overall idea of using a game.

One element that has been investigated in medical learning settings is competition. In a study conducted by Nevin et al. (2014), medical students were given the opportunity to compete in an optional, web-based medical knowledge competition. In this competition, participants answered multiple choice questions pertaining to internal medicine. Participation was voluntary; participants had the ability to earn badges, and there was a dynamically changing leaderboard. When participants were questioned about their experience in subsequent focus groups, the students reported that the leaderboard was their most important motivator (Nevin et al., 2014). Like many studies of games in learning, the focus in data analyses was on acceptance (and in this case, attrition), and there were no measures of objective learning or comparisons that could be used to argue that games enhanced learning beyond business as usual approaches.

A study by Van Nuland et al. (2015) also tested the role of organized competition. Volunteers from an undergraduate functional anatomy course engaged in an online tournament either early or later in the term. In the tournament, participants competed by answering a series of multiple choice or matching questions on the topic of anatomy. Following the competition, students who participated in the activity reported high levels of motivation. The group that competed early in the term, but not those competing later in the term, showed higher course grades than other students in the class who did not participate. Although this study is laudable in using objective measures to compare the gaming groups to non-gaming learners, as in the study by Nevin et al., it is hard to interpret the generality of results given the self-selection of participants. Competition in particular may attract volunteers with greater confidence in their learning of particular subject matter, or gaming may actually energize learning in a subset of individuals (16.7% of the class volunteered in this study). Engagement in the competition also led to more time on task for participants. Interestingly, the authors characterized the intervention as game-like due to competition but attributed the learning gains to the testing effect. Competition and testing can be related, especially in that competition would seem necessarily to involve testing, but they are conceptually separable factors, and testing can be, and often is, implemented without competition.

Perhaps because it has been difficult to show clear learning gains from games, motivational measures are more often reported (Hamari et al., 2014). Here too, the particular elements responsible for group differences may be hard to pin down. Also, enhanced motivation and acceptance presumably have value in learning contexts due to their potential for increasing learning. Studies that relate motivational measures to learning outcomes are likely to be most valuable. Even apart from objective learning measures, it also seems wise to include objective data on motivation such as increased time on task or some other objective measure of participation or persistence, as has been done in some studies, rather than rely exclusively on self-report measures.

Recognition of levels of progress is another potentially valuable element often used in games or gamification. In the same study mentioned above by Nevin et al. (2014), they reported that students who earned badges throughout the competition went on to answer more questions in the activity than those who did not earn badges. That indications of progress might be helpful in increasing motivation and learning is consistent with many different psychological and educational ideas, including reinforcement theory, mastery learning, and goal-directed behavior.

As has been often recognized with the use of standardized tests in schools, it may be important for instructional designers to seek to avoid having the game, rather than the learning content, become the goal. Ng and Bereiter (1991) distinguished three kinds of goals in self-directed learning: task-completion goals, instructional goals, and personal knowledge-building goals. In the learning of basic computer programming, they found that knowledge-building goals, relating to personal interests in learning, were the least frequently expressed, but that students who did express them showed superior learning outcomes in several respects.

Recognition of levels of progress as a learning element readily connects to all three of these goal types. Students' motivation to satisfy the instructional goals (i.e., perform well in the game or on a test) or the task-completion goals, however, may actually compete with knowledge-building (Ng & Bereiter, 1991). We have seen this effect in sequences of trials in adaptive learning where the task and instructional goals of getting to the next trial or finishing a module may cause a learner to speed through and miss valuable trial feedback. A separate issue is that progress indicators may backfire if a learning task is too hard, as participation for long periods without obvious progress undercuts motivation. On balance, however, levels of progress indications seem to have strong potential to maintain motivation. It is worth careful effort, however, to align task and instructional goals with the deeper learning objectives, as game elements, such as speeded response and competition to finish quickly, can act at cross purposes to learning goals.

In STEM and social science subjects, games for primary and secondary school students have often been built around fictional stories, where progress toward the treasure or defeating the space aliens occurs through answering questions or other challenges involving learning items. These approaches have not often been used in medical learning, where the commitment and motivation of learners tends to be higher and more consistent.

5.4.1 Gaming Prospects and Pitfalls

Although more successful versions of learning games could be developed, there are sound scientific reasons that suggest serious limitations on how much games might ultimately enhance meaningful learning. Put another way, these are concerns that identify issues that would have to be addressed or circumvented to derive greater learning benefits from games.

One well-known concern has been called the *bells and whistles* problem (e.g., Moreno & Mayer, 2000). Since the emergence of enhanced graphics, sound, and animation capabilities in ordinary computers, there has been widespread enthusiasm for adding multimedia features, story lines, scenery, characters, sounds, animations, etc. as features in learning software and games in particular. These additions often fail to produce learning benefits and may instead lead to poorer performance. A deep reason is that such additions almost always add extraneous cognitive load (Chandler & Sweller, 1991), which can impair learning of relevant material. Extraneous cognitive load refers to information or inputs that attract processing resources but are not inherently part of the learning content. This kind of load contrasts with intrinsic and/or germane load, which relate to the learning content and associated problem solving efforts (Van Merriënboer & Sweller, 2009); these are relevant to the targeted learning content, and coping with increased load of these types may often be an important factor in learning, as in the “desirable difficulties” framework (Bjork & Bjork, 2006, 2011).

A related problem is what we call the *spoonful of sugar* problem. Embedding mathematics or science learning in a game has often been done on the assumption that the game will be more fun than the actual content of these disciplines. Thus, the engaging storyline, action, or adventure are the “sugar,” whereas the mathematics or science can be slipped into the mix as the medicine, and all will be well. The problem with this reasoning is that students tend to learn the story or game but not so much the math or science (Young et al., 2012). An increase in time and attention will not benefit learning goals if it is primarily dedicated to game elements that are not intrinsically related to the intended academic content. This is not an incidental finding; it is exactly what would be expected on the grounds of limited capacity in human information processing and the extraneous load presented by a story or game. Yet another difficulty is that in any learning domain, some of the material is more amenable to gamification than other parts. Thus, much research on games and learning has essentially “cherry-picked” parts of learning that seem suitable for a game; as a result, there are numerous games in which students advance by correctly doing arithmetic but few that teach quantum mechanics. It remains unclear how general a strategy gamification might be.

Both the bells and whistles problem and the spoonful of sugar problems are related to what has been called the problem or paradox of *seductive details*. Seductive details refer to interesting pieces of information, stories, or images that are only tangentially related to the topic being learned, but are not oriented toward the specific learning objectives (Harp & Mayer, 1998). Such details are added most often in an attempt to make the topic more interesting and relatable for students. However, inclusion of such irrelevant information has been found to actually disrupt learning (Rey, 2012). Moreno and Mayer (2000) discuss seductive details in relation to what they call the *coherency principle*—that multimedia presentations should use minimal extraneous words or images.

Finally, efforts to improve motivation and engagement using games seem to have made little contact with an extensive literature in animal and human behavior about explicit rewards in learning (e.g., Hidi, 2015; Schwartz, 1982). The explicit rewards in a game may lead to short-term engagement but may actually undermine subsequent interest, participation, and learning in that domain when no explicit rewards are offered. We know of no studies that have actually examined long-term effects of game-based learning in this regard.

Medical learning may be in something of a privileged position regarding some of these factors. Medical learners are usually self-selected, highly intelligent, and strongly motivated. Couching learning in a story about fighting off space aliens is usually unnecessary. And competition may be less off-putting than it would be to many students in middle school mathematics. Thus, having a game-like flow, levels of progress, and even competition may have useful applications. These factors may increase motivation, attention, or time on task, and, as mentioned, competition almost always incorporates benefits of the testing effect. Not much research has attempted to disentangle separable effects of these potentially relevant factors.

On balance, continued efforts may lead to development of more valuable gamification efforts. Crucial in such developments will be use of objective measures in learning, as well as avoidance of clear pitfalls of extraneous cognitive load. Most of all, efforts to identify and test rigorously specific game-like elements will have the clearest implications for generalizable efforts.

5.5 Perceptual and Adaptive Learning in Medical Domains

In this section, we consider applications of perceptual and adaptive learning in medical domains. The work in this section focuses on our program of research, which has combined perceptual learning and adaptive learning in efforts to accelerate expertise in a variety of medical learning domains. For brevity, we give some of the general aims and characteristics of PALMs, and we elaborate these in describing PALMs in electrocardiography.

This combination has wide application, as much of medical diagnosis is based upon perceiving and interpreting patterns, be they the details of the squiggles in a 12-lead electrocardiogram (ECG), the sounds transmitted via stethoscope of a beating heart, or specific features in patients' descriptions of why they are consulting a physician. Abilities to pick up the relevant features embedded in each of these, and other, clinical recordings and presentations are not readily mastered but require years of experience, extending from medical school through residency and even throughout the early years of clinical practice. The fact that many patients seek out clinicians who have "been around awhile" rather than newly minted physicians is commonplace, but revealing. Recently trained physicians may have up-to-date factual knowledge and training in the latest techniques, but hard-won expertise in pattern recognition is found in more experienced practitioners.

Although most applications have involved classification of complex clinical displays and presentations, we discuss at the end applications of the core concepts to two other areas in medical learning: factual and conceptual learning that may be greatly improved by adaptive learning methods such as ARTS and higher-order pattern recognition in diagnosis integrating across multiple sources of information. Some work indicates that adaptive and perceptual learning methods also have great potential to enhance these aspects of learning.

5.5.1 General Aims of Perceptual-Adaptive Learning Modules (PALMs)

Why are pattern recognition and diagnostic skills at various levels of medical practice so elusive? Why do they take so long to develop? These questions actually have answers. In fact, there are two related answers. First, as we have already mentioned, the relevant mechanisms involved in these attainments are not much engaged by typical declarative and procedural instructional methods. Much of the relevant learning cannot be well verbalized, and seeing a couple of examples in a textbook or lecture does little to advance perceptual learning.

The second, related, answer is that until recently there has not been much in the way of instructional methods that effectively target pattern recognition. The relevant perceptual learning of classifications, distinguishing features, and the seeing of complex relationships develops gradually and implicitly through encounters with many examples, efforts to classify them (in other words, having a task), and receiving feedback (Kellman & Garrigan, 2009). The instances in any learning domain also need to cover the space in important ways. For a set of diagnostic categories in a domain, for example, training must include ample instances that allow learning mechanisms to sort diagnostic properties that determine membership in one category vs. another, and there must be ample exemplars that show irrelevant variation that spans both categories. Incidental or spurious properties are the bane of effective perceptual category learning. Internship, residency, and actual practice do much to advance the relevant skills, but even there, some phenomena appear rarely, some may be incidentally correlated with other characteristics, and the overall flow of encounters across a learning domain is haphazard.

PALMs were developed to address these problems, based on an emerging understanding of mechanisms of perceptual learning, its role in expertise, and the conditions needed to optimize it. The goals include accelerating the growth of expertise and making learning more comprehensive. The adaptive features of PALMs leverage a number of now well-established learning principles, such as testing and spacing, and use objective data both to guide learning and to certify mastery. The available evidence (see below) suggests that the combined approach can accelerate expertise and produce more comprehensive learning in a variety of domains.

Across various domains, PALMs use the same core features. Trials (learning events) in PALMs consist of classification episodes in which one or more displays are presented and learners must make a response, such as classifying a display, comparing displays, or distinguishing some feature or property. Accuracy and response times are used for spacing, to determine progress and eventually mastery, based on both accuracy and fluency. Feedback occurs at the trial level, after blocks of trials, and in terms of levels of progress indicated periodically. Usually, mastery criteria are implemented, such that learning continues until all learning categories have been mastered. Progress tracking spans sessions and allows learns to return and continue learning where they left off. In research investigations of PALMs with real medical learners (medical students, residents, or practitioners), we usually employ designs in which assessments are administered at pre-test and post-test and in a delayed post-test, administered weeks, or sometimes months, later. Active control groups are sometimes used, but comparisons to non-users (business-as-usual control) and comparisons of PALM users to more advanced practitioners (e.g., comparing 3rd-year medical students to 2nd-year residents) have often been a more practical methodology than subjecting a separate group of learners to some other intervention.

We have created and investigated PALMs in the areas of anatomy, cardiology, dermatology, histopathology, obstetrics and gynecology, ophthalmology, radiology, and ultrasound. Details for many of these interventions have been published or are under review (Krasne et al., 2013, 2020; Rimoin et al., 2015; Romito et al., 2016; Slaught et al., 2021).

5.5.2 Electrocardiography PALMs

Learning to discriminate pathologies underlying the patterns produced in 12-lead electrocardiograph recordings used in the PALMs relies very heavily on perceptual learning, since there are large variations among individual patients for both normal patterns and in features associated with a given pathology, and relevant features are embedded in varying degrees of “noise.” Thus, beyond knowing the names of the categories and the general features associated with each, mastery primarily involves more selective extraction of subtle features and detection of important relations, as well as quicker and more effortless performance. Some of the relevant complexity is illustrated by the electrocardiogram (ECG) traces in Fig. 5.1.

Here is a bit of background for understanding how to view a 12-lead ECG, such as those shown in Fig. 5.1. Each of the 12 panels display 2.5 s traces of electrical signals arising from different combinations of electrodes placed on the body in a specified pattern. The vertical axis of the traces displays the electrical signal in millivolts, and the horizontal axis displays time in seconds. Three electrical signal traces are displayed at a time for 2.5 s before switching to another set of three traces. The labels on the traces are standard, referring to the particular combination of electrodes from which the recording is derived.

ECGs from two different normal patients are shown in Fig. 5.1a, b, whereas Fig. 5.1c, d displays two patients with an acutely occurring type of heart attack

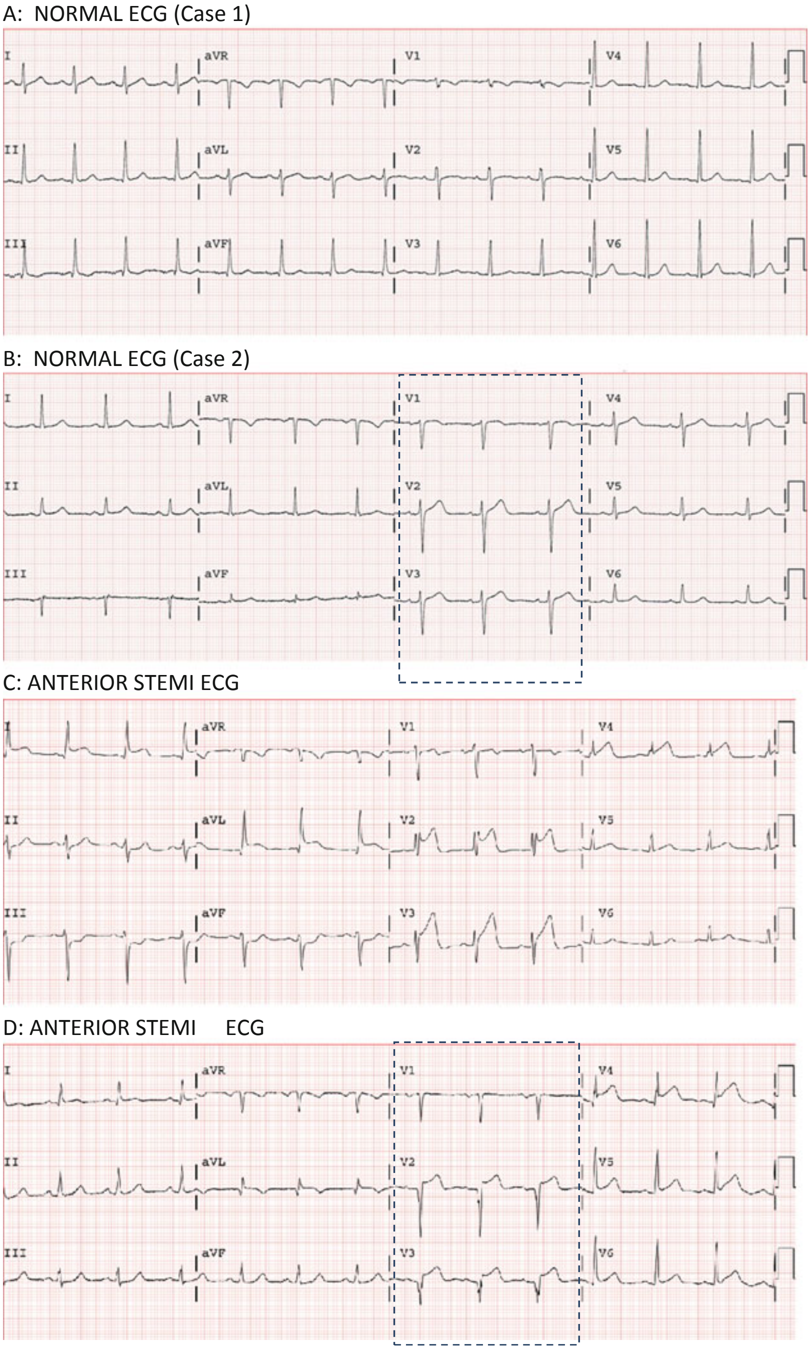

Fig. 5.1 Examples of 12-lead echocardiograms (ECGs), as used in the ECG Morphology PALM. (**a**) and (**b**) show two different normal cases. (**c**) and (**d**) show two different patients with anterior STEMI. Note the complexity of the information displayed and the variation for cases within each

arising in the anterior portion of the left ventricle, referred to as an anterior ST-elevated myocardial infarction (STEMI).

These records illustrate two of the problems inherent in classifying ECGs as well as in other medical classification tasks: Within any given diagnostic category, there is a large range of variation (as can be seen in Fig. 5.1a vs. b and Fig. 5.1c vs. d); and between diagnostic categories, there is a large overlap in appearance (Fig. 5.1b vs. d). Of particular relevance in distinguishing the type of heart attack shown are the traces within the dashed rectangles superimposed on each recording. Although these traces appear quite similar, a key difference in determining that Fig. 5.1d, but not b, represents an acutely occurring heart attack is that the upswing following the large dip in traces V2 and V3 of d overshoots each trace's baseline by at least two of the small horizontal voltage lines prior to bending, whereas those in the equivalent traces of Fig. 5.1b begin bending when they reach the baseline. Diagnoses of a large number of cardiac disorders are made based on small differences such as these or even less obvious ones, occurring in one or more traces, and expertise in picking out such differences requires many years of experience.

Because interpreting ECGs is difficult and critical to making rapid and accurate diagnoses of cardiac problems, we developed PALMs to improve the early mastery of this clinical tool. Most extensively, we studied, over a broad range of educational levels and retention times, outcomes associated with an *ECG Morphology PALM* that focuses on the shapes of the electrical signals, rather than their rhythmicity. We use this specific PALM to illustrate some features of medically related PALMs and their performance. Further detail and all of the data can be found in Krasne et al. (2020).

In our experiments on the ECG Morphology PALM, we evaluated its effectiveness by comparing performance on pre-test assessment prior to PALM training and on a post-test following PALM training. The ECG exemplars used in assessments had not been seen previously and were different for pre- and post-tests. Effects of the PALM intervention on classifying 12-lead ECGs into 15 different diagnostic categories were examined for third- and fourth-year medical students (MS3 and MS4, respectively) and emergency medicine residents in their first, second, and third years (EMR1, EMR2, and EMR3). Individual progress through the PALM was tracked for each individual in terms of accuracy and response times for instances of each diagnostic category. Mastery criteria including accuracy, speed, and success across spaced delays were applied to each diagnostic category, such that learning in the PALM by a given learner continued until mastery was attained for each of the learning categories. The durations of participation required to reach mastery varied by educational level, ranging from 88 ± 32 min for MS3 students to 46 ± 24 min for EMR3 residents.

Figure 5.2 shows results of the ECG Morphology PALM for a range of participants from MS3s to EMR3s based on pre-test and post-test performance on an

Fig. 5.1 (continued) category. Becoming able to accurately choose a diagnostic category from these displays requires perceptual learning to select relevant information, detect subtle features, and extract important relations, as well as ignore irrelevant variations (see text)

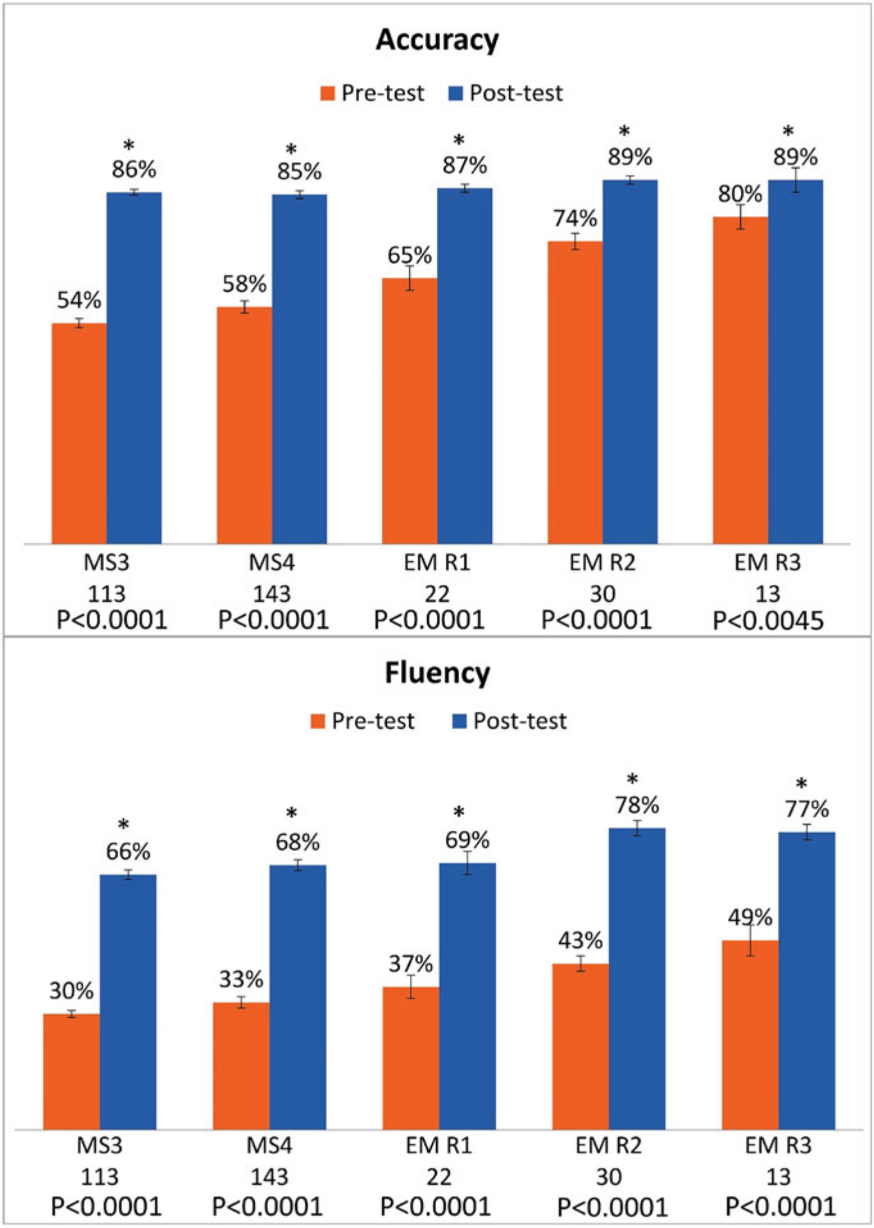

Fig. 5.2 Effect of the ECG Morphology PALM on ECG interpretation skill. *Upper panel*: Percent accurate classification shown as a function of educational level for ECG interpretation before (orange bars or gray in grayscale version) and after (blue bars or black in grayscale version) completing the ECG Morphology PALM for the first time. MS3 and MS4 were medical students in years 3 and 4; EM R1, EM R2, and EM R3 were emergency medicine residents in years 1, 2, and 3. *Lower panel*: Fluency, defined as accurate responding on assessment ECGs with a response time ≤15 s is shown as a function of educational level. For both panels, numbers of participants are

assessment of ECG interpretation. We used two primary measures: *accuracy*, defined as proportion of correct responses made within the 30 s allotted time period for each trial, and *fluency*, which measures the fraction of responses that were both accurate and completed within 15 s. By examining responses made within a shorter target time, the fluency measure aims to capture changes in perceptual extraction of information and progress toward automaticity rather than simply the ability to consciously analyze the ECG.

Initial (pre-test) performance reflected educational level. Performance following the PALM training greatly improved and was much more equalized across the different training levels. Besides the robust improvement for participants at all stages of medical training, the comparisons across educational levels provide evidence that PALM training leads to acceleration of proficiency in these skills relative to conventional training. Note that 3rd- and 4th-year medical students at the end of PALM use achieved levels of performance that exceeded pre-test performance of emergency medicine residents in years 1, 2, and 3. Interpreting ECGs is an important part of emergency medicine training, and residency provides considerable exposure to ECGs and opportunities to develop these skills. The monotonic increase in pre-test scores for EM1 through EM3 suggests that proficiency does increase across years of residency. By comparison, however, only 1–2 h of the ECG Morphology PALM produced much larger gains, such that PALM-trained medical students not only exceeded residents' pre-test accuracy in classifying ECGs, but did not differ much from residents' post-test accuracy scores.

In the fluency data, some of these patterns were even more pronounced. There, 3rd- and 4th-year medical students at post-test substantially outperformed residents in any year who had not undergone PALM training. It is also clear, however, that residents' fluency benefitted greatly from the PALM, with EM2s and EM3s going from 43 and 49% to 78 and 77%, respectively. As with the accuracy data, there was a clear increase in pre-test fluency across years of residency, suggesting that conventional training does improve these skills, but these gains were somewhat subtle (37–49% from EM1 to EM3) compared to the gains from the 1–2 h PALM (which led to gains in fluency of 32, 35, and 28 percentage points for the EM1, EM2, and EM3 groups, respectively). The gains in both fluency and accuracy suggest that this adaptive perceptual learning intervention effectively addressed both discovery and fluency aspects of perceptual learning (Kellman, 2002). The especially large gains in fluency suggest that PALM training may specifically impact aspects of learning that are relatively poorly addressed through conventional instruction or apprenticeship (although it is also possible that fluency gains appear to exceed accuracy gains due to a ceiling effect in the accuracy data).

Fig. 5.2 (continued) shown below the training level on the abscissa. For all groups and measurements, the post-test performance was statistically better than the pre-test performance with $P < 0.0001$ for all pre-/post-test comparisons except for R3 Accuracy, for which $P = 0.0045$

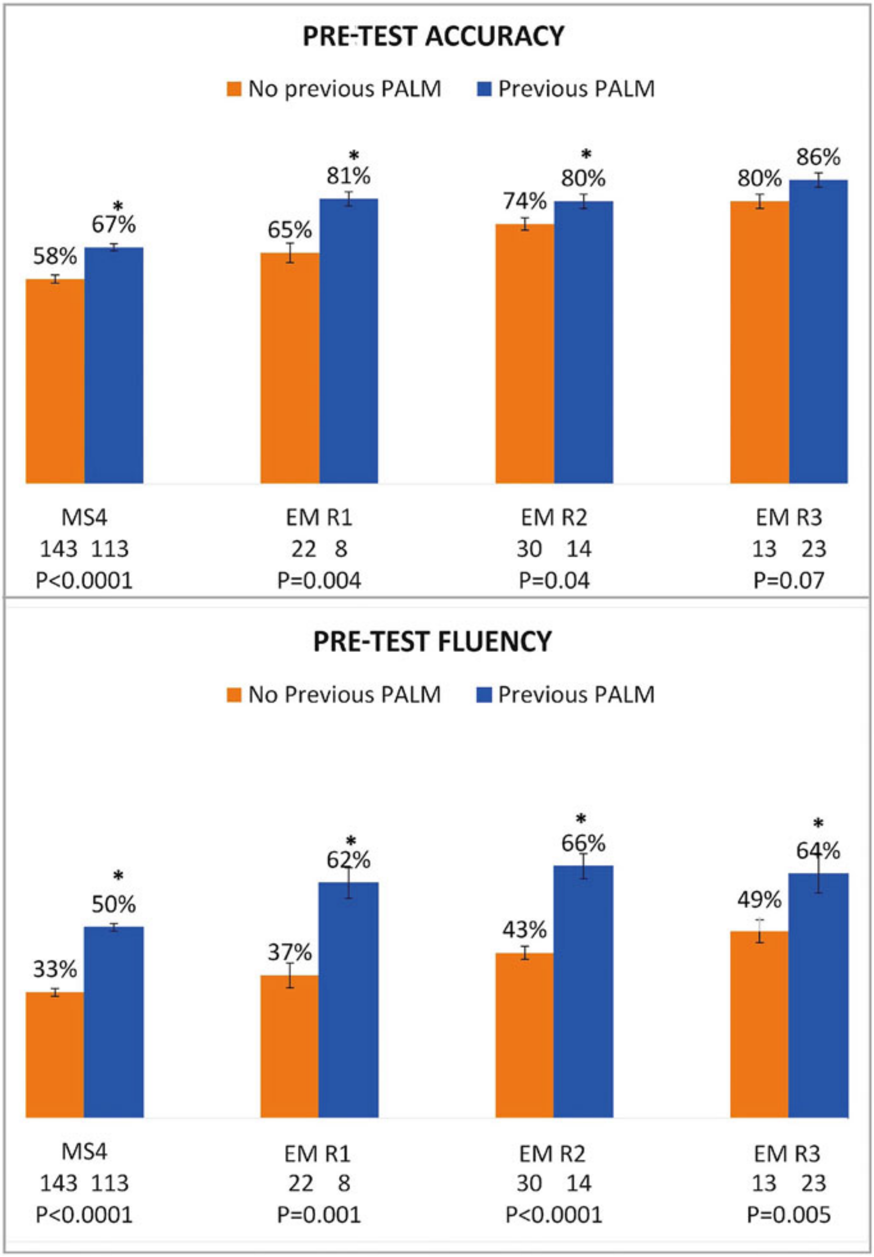

Fig. 5.3 Delayed test effects of ECG Morphology PALM training on ECG assessment performance. Left-hand bars in each pair (orange in color version of the figure; gray in grayscale version) represent trainees at several educational levels who had no previous ECG PALM experience. Right-hand bars in each pair (blue in color version of the figure; black in grayscale version) represent current performance on the same assessment for trainees at the same educational level who had completed the ECG PALM approximately 1 year previously. Numbers of participants are shown

We were able to assess some of the long-term effects of completion of the ECG Morphology PALM by testing a year later participants of equivalent educational level who had or had not previously used the PALM. These data can be seen in Fig. 5.3. (The data are labeled as "pre-test" scores because these learners went on to complete the PALM afterward.) At each level, participants who had completed the 1–2 h PALM a year or so earlier outperformed peers who had not done so. With the exception of the accuracy performance of R3 residents (for which $P = 0.07$), both accuracy and fluency scores for trainees who had completed the PALM approximately 1 year earlier were statistically significantly higher than those for trainees at the same educational level but with no previous ECG PALM training. Effect sizes were large for all fluency comparisons and were medium to large for accuracy comparisons. Another comparison that can be made from the data of Fig. 5.3 is the relative effectiveness of the PALM training versus that of the clinical experience of the emergency medicine residents during the previous year. The accuracy performance of residents completing ECG PALM training 1 year previously is comparable to or greater than that of residents 1 year more advanced who had spent the previous year in the clinic but without PALM training (compare right hand [black or orange] bars for MS4, EMR1, and EMR2 with left hand [gray or blue] bars for EMR1, EMR2, and EMR3, respectively). In comparing fluency performance, the effectiveness

of the PALM is even greater than that of the additional year of clinical experience of residents in the next highest year of training. These results indicate substantial lasting gains from PALM training. We know of no other learning interventions that have been shown to produce meaningful and enduring learning gains a year later from a 1–2 h activity.

5.5.3 PALMs in Other Medical Learning Domains

The learning principles and algorithms in PALMs have broad applicability, and they have proven to be useful in many domains, including mathematics and STEM learning (Kellman et al., 2010; Mettler, Massey, et al., 2020; Unuma et al., 2016), chemistry (Mettler, El-Ashmawy, et al., 2020), music (Bufford et al., 2016) and aviation training (Kellman & Kaiser, 1994). As mentioned earlier, we have developed and tested PALMs to improve the learning of difficult perceptual classifications in a variety of other medical learning domains, including histopathology (Krasne et al., 2013), dermatology (Rimoin et al., 2015; Slaught et al., 2021), and trans-

Fig. 5.3 (continued) below the educational level on the abscissa along with P values (two-sided) for assessment performance differences between those with and those without the ECG PALM experience the previous year. For accuracy, effect sizes were $d = 0.7$, 1.7, 0.5, and 0.7, and for fluency, they were $d = 1.3$, 1.9, 1.8, and 0.8 for MS4, R1, R2, and R3 trainees, respectively esophageal echocardiography (Romito et al., 2016). Uniformly, we have found that PALM training appears to produce large, significant, durable improvements in performance for a variety of clinical tests and presentations. PALMs or elements of PALMS are increasingly being applied by others interested in improving medical learning (e.g., Guerlain et al., 2004; Sha et al., 2020; Perham et al., 2015), have been evaluated in a systematic review (Guegan et al., 2020), and have been the subject of favorable commentaries in medical journals (e.g., Evered, 2018; Waite et al., 2020; Weller, 2016).

In addition, surveys that we implemented early on in conjunction with our first medical PALM applications, those for training basic dermatology features and histopathology processes, demonstrated strong support for PALM implementation in the curriculum, with MS1 students averaging 4.5/5 on a Likert scale evaluating whether there should be comparable modules for other areas of the curriculum and MS2 students averaging 4.6/5 for the same issue (Krasne et al., 2013; Rimoin et al., 2015).

Finally, the gamification features discussed earlier in this chapter were clearly recognized and influenced the positive response to the PALM modules, with students spontaneously including characterizations such as the following:

> . . . really nice...and made it seem like a game, which I enjoyed and wish we had more of.
> . . . lots of examples, game format made it fun
> The derm and histo 'games' were extremely helpful . . .
> . . . it was fun and stress-free, more like a game as opposed to a test.
> The game-type set up made the exercise fun
> —repetition—use of images—"game" format—diagnosis—interactive-ness of the program

5.5.3.1 Factual Learning, Simulation, and Higher-Order Pattern Recognition in PALMs

Although we have emphasized applications to perceptual classifications, PALMs also have considerable potential to enhance other kinds of learning. The adaptive learning algorithms in PALMs are well-suited to learning of factual material, and evidence suggests that the ARTS system in particular offers advantages over other spacing schemes (Mettler et al., 2016). Many aspects of medical learning involve substantial learning of facts, concepts, and procedures, all of which can be enhanced by the testing, spacing, and mastery components of ARTS. Yet another difficult but common learning challenge involves integration of various signs, symptoms, patient characteristics, and test results in medical diagnosis. This can be considered a higher-order pattern recognition problem, which, although it may involve perceptual components, is not primarily perceptual. The formal similarity to perceptual learning, in terms of discovering distinguishing features and important relations, and acquiring fluency, however, has suggested that a PALM approach may be capable of accelerating learning in diagnostic domains.

Some work has tested this hypothesis. Kellman and Mettler (2011) studied PALMs in a US Army project aimed at teaching information and diagnostic skills at different levels. The focus was on chief complaints that might present among military personnel and their families at military installations. Adaptive factual learning modules (FLMs) were developed to teach the basics of specific medical conditions that might lead to a particular chief complaint (e.g., meningitis, systematic viral infection, subarachnoid hemorrhage, migraine, or tension headache for the complaint of headache). Integrated factual learning modules (IFLMs) mixed learning across medical conditions with the aim of sharpening long-term retention and understanding of distinctions. Finally, cognitive task modules (CTMs) presented novel cases that trained both the procedural aspects of treatment and investigatory steps related to different specific conditions as well as differential diagnosis based on sets of characteristics, signs, symptoms, and test results. Modules at all levels used the ARTS system for adaptive learning and mastery. Undergraduate participants with health-related career interests served as participants. Results showed substantial advantages of adaptive learning modules over conventional study materials, and they also showed that the sequence from FLMs to CTMs added considerably to mastery and later diagnostic accuracy. PALMs have also shown great utility in anatomy learning, which is a combination of factual and perceptual learning. Suites of anatomy learning focused on specific body areas or for different purposes (e.g., surgery vs. physical therapy training) have been developed by Insight Learning Technology, Inc. and Primal Pictures, Inc., a leading distributor of high-quality anatomy graphics online.

In another project (Benharash et al., 2015), PALMs were applied to simulation in trauma care. The goal of the project was to develop an integrated simulation and training platform based on perceptual and adaptive learning principles, focused on training combat medics in the factual and conceptual underpinnings, diagnosis, and emergency treatment in hemorrhage control. The trauma care simulator displayed vital signs of a patient, presented other information, and could present queries or cases. Training interleaved several levels of interactive learning. Recognition trials trained learners to fluently recognize important indications, such as a vital sign out of normal limits. Some examples included blood pressure, respiration rate, and cardiac arrhythmias. Diagnostic trials involved higher-order pattern recognition from combined simulator and case information inputs. Examples included hypovolemic shock, neurogenic shock, tension pneumothorax, cardiac tamponade, and traumatic brain injury. Treatment trials involved administration of medications, electric shock, airway interventions, etc. and assessment of their consequences in the simulator. Some examples included pressure or tourniquet for hemorrhage, shock for ventricular tachyarrhythmias, morphine administration, and resolution of airway issues through intubation, O_2 administration, or cricoidotomy. Finally, sets of factual learning points were included to teach the MARCH algorithm, which gives an order of priority for initial assessment and treatment for trauma victims (M, massive bleeding; A, airway; R, respirations; C, circulation; and H, head).

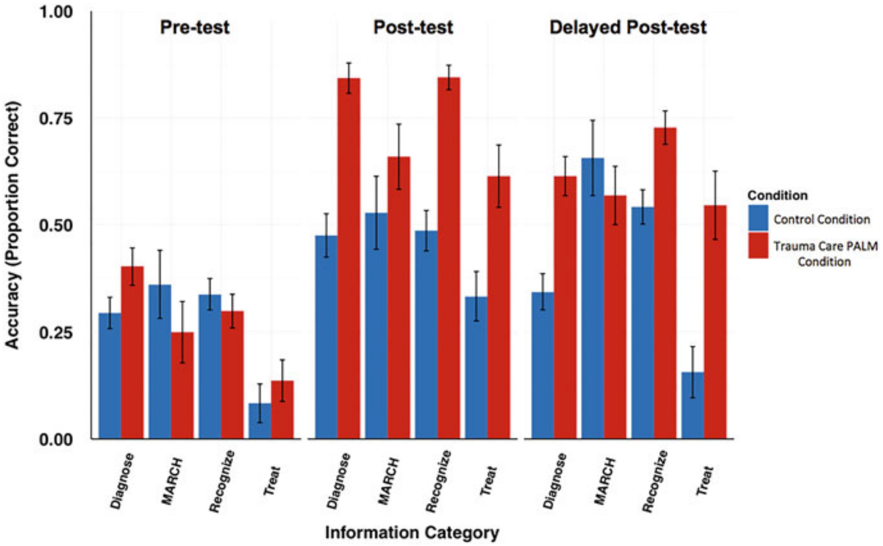

Fig. 5.4 Assessment data by Trial Type and Assessment Phase for trauma care training. Blue bars (black in grayscale version) in each pair show control group data. Red bars (gray in grayscale version) show data from the Trauma Care Simulation PALM. Pre-test, post-test, and delayed post-test data are shown for the four types of learning trials: diagnostic trials, MARCH algorithm trials, recognition trials, and treatment trials. Error bars show ± standard error of the mean

One group learned entirely in the Trauma Care Simulator PALM, and a control group learned from actual study materials used in basic military medical training. The learning content was the same in both groups.

Assessments were given half within the simulator environment and half as paper and pencil tests. The Trauma Care Simulation PALM produced substantially better accuracy in performance on both types of tests, both in immediate post-tests and in a delayed post-test 1 week later. Figure 5.4 shows the patterns of data for the simulation-based assessments. The four learning areas (diagnosis, MARCH algorithm, recognition, and treatment) are shown from left to right for the pre-test, post-test, and delayed post-test, given 1 week later. There are conspicuous advantages for the PALM group in diagnostic accuracy, recognition, and selection of treatment actions. These are evident at immediate post-test and also after a delay. For the MARCH algorithm, the group trained using simulation and adaptive learning showed a trend toward better learning at immediate post-test, especially given an advantage for the control group at pre-test, but the difference in learning gains was not different between groups at delayed post-test. The treatment category appeared to be the hardest to learn, in terms of immediate post-test scores, but what was learned declined very little between immediate and delayed post-tests for the simulation group, especially in comparison with the control group. This might indicate greater incorporation of pattern recognition components when the learning occurred in the PALM format. Oddly, the PALM group outperformed the control group on MARCH algorithm content, which might be considered the most purely declarative and procedural learning content in the study, when tested in paper and pencil format (as well as in all of the other learning categories).

These results suggest potential for much broader application of perceptual and adaptive learning technologies in both factual learning in medicine and in more complex settings, including simulation, integration of information in diagnosis, and the application and assessment of treatments.

5.6 Conclusion

Instruction and learning in medical domains can be improved by application of findings from the science of learning and by advances in technology. For new approaches or enhancements to be effective requires fruitful combinations of these. We have considered a number of current and emerging ideas with already realized or potential value in medical education. Scientific research and technological implementations involving perceptual learning add value both by broadening our conceptions of what learning includes and how it can be systematically advanced. Advances in digital technology and principles of learning make possible adaptive methods that can personalize learning and make it more efficient, as well as track progress and ensure more comprehensive mastery. Efforts in these areas, basic and applied, are also expanding ideas of mastery to include fluency, automaticity, and transfer to new situations. Adaptive learning technologies have powerful applications when combined with perceptual learning, as we have illustrated, and they also have compelling potential to improve factual and procedural learning. Games and gamification may be best considered in terms of useful elements; there is no magic for learning that accrues just because something is a game. Specific components of gaming may have value in learning, especially those that have been or could be subjected to experimental work using objective measures of learning that can test their effectiveness and reveal best practices. Promising elements include progress tracking, feedback, and potential effects on motivation, whereas challenges include avoidance of distracting seductive details, extraneous cognitive load, and fostering intrinsic motivation that might persist outside of the gamified learning environment. The continuing emergence of findings from scientific research on learning and new technologies in which they can be instantiated promises a bright future for evidence-based advances in instruction, but these will also require continuing, committed, and collaborative efforts from domain experts and learning scientists to devise, test, and implement new methods that truly add value.

References

Abbott, E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs: General and Applied, 11*(1), 159–177.

Atkinson, R. C. (1976). Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and instruction*. Wiley.

Benharash, P., Older, R., Mettler, E., Genovese, B., Frank, P., Dutson, E., & Kellman, P. (2015). Perceptual and adaptive learning and simulation in training combat medics in trauma care. *UCLA human perception laboratory technical report to ONR cognitive science of learning program review*, Sept., 2016.

Benzil, D. L. (2018). See one, do one, teach one. *AANS Neurosurgeon, 27*(3).

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society, 2*, 59–68.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Lawrence Erlbaum.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes, 2*, 35–67.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L.-G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 109–133). Psychology Press.

Bondie, R. S., Dahnke, C., & Zusho, A. (2019). How does changing "one-size-fits-all" to differentiated instruction affect teaching? *Review of Research in Education, 43*, 336–362. https://doi.org/10.3102/0091732X18821130

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. National Academies Press.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6*, 345–375.

Bufford, C. A., Thai, K. P., Ho, J., Xiong, C., Hines, C. A., & Kellman, P. J. (2016). Perceptual learning of abstract musical patterns: Recognizing composer style. In T. Zanto, T. Fujioka, P. Janata, J. Johnson, J. Berger, J. Slater, & C. Chafe (Eds.), *Proceedings of the 14th annual international conference for music perception and cognition* (pp. 8–12).

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369–378.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*(6), 632–642. https://doi.org/10.3758/bf03202713. PMID 1435266. S2CID 15893469.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

De Groot, A. D. (1965). *Thought and choice in chess*. Mouton.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15).

Engeström, Y., Miettinen, R., & Punamäki, R. (Eds.). (1999). *Perspectives on activity theory*. Cambridge University Press.

Evered, A. (2018). Perceptual and adaptive learning modules and their potential to transform cytology training programmes. *Cytopathology, 29*(4), 371–374.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton-Mifflin.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*(2), 95–112.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 86–112.

Griff, E. R., & Matter, S. F. (2013). Evaluation of an adaptive online learning system. *British Journal of Educational Technology, 44*(1), 170–176.

Guegan, S., Steichen, O., & Soriace, A. (2020). Literature review of perceptual learning modules in medical education: What can we conclude regarding dermatology? *Annales de Dermatologie et de Vénéréologie*. Available online 9 July 2020. https://doi.org/10.1016/j.annder.2020.01.023

Guerlain, S., Green, K. B., LaFollette, M., Mersch, T. C., Mitchell, B. A., Poole, G. R., Calland, J. F., Lv J.,, & Chekan, E. G. (2004). Improving surgical pattern recognition through repetitive viewing of video clips. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 34*(6), 699–707.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?—A literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences* (pp. 3025–3034). IEEE.

Hamnvik, O. R., Alford, D. P., Ryan, C. T., Hardesty, I. T., & Drazen, J. M. (2019). NEJM knowledge+ pain management and opioids—A new adaptive learning module. *New England Journal of Medicine, 380*, 1576–1577.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414.

Healy, M. G., Phitayakorn, R., Axelsson, C. G., O'Rourke, M., Duhigg, L., & Gray, B. A. (2019, February 28). *Novel adaptive e-learning product helping general internists pass their ABIM IM-MOC exam*. Presented at the Society for Academic Continuing Medical Education annual meeting, Charleston, SC.

Hidi, S. (2015). Revisiting the role of rewards in motivation and learning: Implications of neuroscientific research. *Educational Psychology Review, 28*(1), 61–93. https://doi.org/10.1007/s10648-015-9307-5

Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2*, 153–160.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97–103.

Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704–719.

Kellman, P. J. (2002). Perceptual learning. In H. Pashler & R. Gallistel (Eds.), *Stevens' handbook of experimental psychology: Learning, motivation, and emotion*. Wiley.

Kellman, P. J. (2006). *System and method for adaptive learning*. Inventor: P. Kellman (US PTO #7052277, awarded 4/2006).

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews, 6*, 53–84.

Kellman, P. J., & Kaiser, M. K. (1994). Perceptual learning modules in flight training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*, 1183–1187.

Kellman, P. J., & Krasne, S. (2018). Accelerating expertise: Perceptual and adaptive learning technology in medical learning. *Medical Teacher, 40*(8), 797–802.

Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. In *Psychology of learning and motivation* (Vol. 58, pp. 117–165). Academic Press.

Kellman, P. J., Massey, C. M., & Son, J. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science (Special Issue on Perceptual Learning), 2*(2), 285–305.

Kellman, P. J., & Mettler, E. M. (2011). *ACAT technology for optimal sick call performance*. UCLA Human Perception Laboratory Technical Report, US Army RDECOM Contract No: W91CRB-09-C-0093.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19*(6), 585–592.

Kornell, N., & Hausmann, H. (2017). Performance bias: Why judgments of learning are not affected by learning. *Memory and Cognition, 45*, 1270–1280.

Krasne, S., Hillman, J. D., Kellman, P. J., & Drake, T. A. (2013). Applying perceptual and adaptive learning techniques for teaching introductory histopathology. *Journal of Pathology Informatics, 4*, 34. https://doi.org/10.4103/2153-3539.123991

Krasne, S., Stevens, C. D., Kellman, P. J., & Niemann, J. T. (2020). Mastering electrocardiogram interpretation skills through a perceptual and adaptive learning module. *Academic Emergency Medicine Education and Training, 5*(*2*), *e10454*. https://doi.org/10.1002/aet2.10454. eCollection 2021 Apr. PMID: 33796803 1–11.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg et al. (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Larsen, D. P., Butler, A. C., & Roediger, H. L., III. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47*(*7*), 674–682.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59*, 277–287.

McMahon, G. T., & Drazen, J. M. (2014). Introducing NEJM Knowledge+ and its adaptive personalized learning. *New England Journal of Medicine, 370*(17), 1648–1649.

Mettler, E., El-Ashmawy, A. K., Massey, C. M., & Kellman, P. J. (2020). Adaptive vs. fixed spacing of learning items: Evidence from studies of learning and transfer in chemistry education. In *Proceedings of the 42nd annual conference of the Cognitive Science Society* (pp. 1598–1604). Cognitive Science Society.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research, 99*, 111–123.

Mettler, E., Massey, C., Burke, T., & Kellman, P. J. (2020). Comparing adaptive and random spacing schedules during learning to mastery criteria. In *Proceedings of the 42nd annual conference of the Cognitive Science Society* (pp. 773–779). Cognitive Science Society.

Mettler, E., Massey, C., & Kellman, P. J. (2011). Improving adaptive learning technology through the use of response times. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2532–2537). Cognitive Science Society.

Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General, 145*(*7*), 897–917.

Moreno, R., & Mayer, R. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(*1*), 117–125.

Nevin, C. R., Westfall, A. O., Rodriguez, J. M., Dempsey, D. M., Cherrington, A., Roy, B., Patel, M., & Willig, J. H. (2014). Gamification as a tool for enhancing graduate medical education. *Postgraduate Medical Journal, 90*(1070), 685–693.

Ng, E., & Bereiter, C. (1991). Three levels of goal orientation in learning. *Journal of the Learning Sciences, 1*(3,4), 243–271.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing instruction and study to improve student learning. IES practice guide. *NCER 2007–2004*. National Center for Education Research.

Perham, N., Evered, A., Walker, D., & Watt, A. (2015). Visual perceptual learning in cytopathology. In T. Heinen (Ed.), *Advances in visual perception* (pp. 203–232). Nova Science.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437–447.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*(3), 216–237.

Rimoin, L., Altieri, L., Craft, N., Krasne, S., & Kellman, P. J. (2015). Training pattern recognition of skin lesion morphology, configuration, and distribution. *Journal of the American Academy of Dermatology, 72*(3), 489–495.

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(*1*), 20–27.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture and Activity, 1*(4), 209–229. https://doi.org/10.1080/10749039409524673

Romito, B. T., Krasne, S., Kellman, P. J., & Dhillon, A. (2016). The impact of a perceptual and adaptive learning module on transoesophageal echocardiography interpretation by anaesthesiology residents. *British Journal of Anaesthesia, 117*(4), 477–481.

Rutledge, C., Walsh, C. M., Swinger, N., Auerbach, M., Castro, D., Dewan, M., Khattab, M., Rake, A., Harwayne-Gidansky, I., Raymond, T., Maa, T., & Chang, T. (2018). Gamification in action: Theoretical and practical considerations for medical educators. *Academic Medicine, 93*(7), 1014–1020.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(*1*), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Perspectives on Psychological Science, 13*(*2*), 146–148.

Schwartz, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General, 111*(1), 23–59.

Sha, L. Z., Toh, Y. N., Remington, R. W., & Jiang, Y. (2020). Perceptual learning in the identification of lung cancer in chest radiographs. *Cognitive Research: Principles and Implications, 5*(4).

Sharma, N., Doherty, I., & Dong, C. (2017). Adaptive learning in medical education: The final piece of technology enhanced learning? *The Ulster Medical Journal, 86*(3), 198.

Slaught, C., Madu, P., Chang, A., Williams, V., Kebaetse, M., Nkomazana, O., Molefe-Baikai, J., Bekele, N., Omech, B., Kellman, P., Krasne, S., & Kovarik, C. (2021). Novel education modules addressing the underrepresentation of skin of color in dermatology training. *Journal of Cutaneous Medicine & Surgery*. Advanced online publication. https://doi.org/10.1177/12034754211035093

Spector, J. M. (2014). Conceptualizing the emerging field of smart learning environments. *Smart Learning Environments, 1*(1), 1–10.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*, 244–253.

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning & Verbal Behavior, 15*(5), 529–536.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175–184.

Unuma, H., Hasegawa, H., & Kellman, P. J. (2016). Perceptual learning facilitates precise mental representations of fractions. *The Journal of Kawamura Gakuen Women's University, 27*(1), 35–49.

Van Merriënboer, J. J. G., & Sweller, J. (2009). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education, 44*(85-93), 16.

Van Nuland, S. E., Roach, V. A., Wilson, T. D., & Belliveau, D. J. (2015). Head to head: The role of academic competition in undergraduate anatomical education. *Anatomical Sciences Education, 8*(5), 404–412.

Waite, S., Farooq, Z., Grigorian, A., Sistrom, C., Kolla, S., Mancuso, A., Martinez-Conde, S., Alexander, R. G., Kantor, A., & Macknik, S. L. (2020). A review of perceptual expertise in radiology: How it develops, how we can test it, and why humans still matter in the era of artificial intelligence. *Academic Radiology, 27*(1), 26–38.

Weller, J. M. (2016). New approaches to the development of expertise in anaesthesia. *British Journal of Anaesthesia, 117*(5), 545–546. https://doi.org/10.1093/bja/aew329

Wiseheart, M., D'Souza, A. A., & Chae, J. (2017). Lack of spacing effects during piano learning. *PLoS One, 12*(8), e0182986.

Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell.

Yardley, S., Teunissen, P. W., & Dornan, T. (2012). Experiential learning: Transforming theory into practice. *Medical Teacher, 34*(2), 161–164. https://doi.org/10.3109/0142159X.2012.643264

Young, M. F., Slota, S. T., Cutter, A., Jalette, G. C., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89. https://doi.org/10.3102/0034654312436980