Published in final edited form as:

J Sci Educ Technol. 2018 December ; 27(6): 566–580. doi:10.1007/s10956-018-9748-y.

Learning Neuroscience with Technology: A Scaffolded, Active Learning Approach

Katrina B. Schleisman^{a,e}, S. Selcen Guzey^{b,c}, Richard Lie^c, Michael Michlin^d, Christopher Desjardins^d, Hazel S. Shackleton^a, August C. Schwerdfeger^a, Martin Michalowski^a, and Janet M. Dubinsky^{e,*}

^aAndamio Games, 111 Third Avenue South, Suite 100, Minneapolis, MN, 55401.

^bPurdue University, Department of Curriculum & Instruction, 4108 Beering Hall, 100 North University Street, West Lafayette, IN, 47907.

^cPurdue University, Department of Biological Sciences, 915 West State Street, West Lafayette, IN, 47907

^dUniversity of Minnesota, Center for Applied Research & Educational Improvement, 1954 Buford Avenue, Suite 425, St. Paul, MN, 55108

^eUniversity of Minnesota, Department of Neuroscience, 6-145 Jackson Hall, 321 Church Street Southeast, Minneapolis, MN, 55455

Abstract

Mobile applications (apps) for learning technical scientific content are becoming increasingly popular in educational settings. Neuroscience is often considered complex and challenging for most students to understand conceptually. iNeuron is a recently developed iOS app that teaches basic neuroscience in the context of a series of scaffolded challenges to create neural circuits and increase understanding of nervous system structure and function. In this study, four different ways to implement the app within a classroom setting were explored. The goal of the study was to determine the app's effectiveness under conditions closely approximating real-world use, and to evaluate whether collaborative play and student-driven navigational features contributed to its effectiveness. Students used the app either individually or in small groups, and used a version with either a fixed or variable learning sequence. Student performance on a pre- and post- neuroscience content assessment was analyzed and compared between students who used the app and a control group receiving standard instruction, and logged app data were analyzed. Significantly greater learning gains were found for all students who used the app compared to control. All four implementation modes were effective in producing student learning gains relative to controls, but did not differ in their effectiveness to one another. In addition, students demonstrated transfer of information learned in one context to another within the app. These results suggest that teacher-led

*Corresponding author. dubin001@umn.edu, Phone: +1 612 625 8447.

Disclosure of conflicts of interest: HSS, ACS, MM are employees of the company that owns the app and may benefit from its sale. KBS has multiple affiliations, and was eventually hired as an employee of the company that owns the app.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

neuroscience instruction can be effectively supported by a scaffolded, technology-based curriculum which can be implemented in multiple ways to enhance student learning.

Keywords

educational games; educational technology; neuroscience education; student learning

The advance of technology use in classrooms has created new opportunities to develop curriculum for difficult-to-teach science subjects. Neuroscience is a key curricular component for anatomy and physiology, psychology, and biology high school courses. However, it is perceived to be a complex topic with few resources for secondary teachers (grades 9–12) beyond college-level textbooks. Some high-quality online resources are available (Chudler & Bergsman 2014; Schotland & Littman 2012), but most are geared towards educating the general public or K-12 students at elementary and middle school grade levels. Furthermore, these resources often target specific, health-related neuroscience topics (e.g., *Mouse Party*, Genetic Science Learning Center, 2013; and *Sowing the Seeds of Neuroscience*, www.neuroseeds.org).

iNeuron was developed to provide a comprehensive introduction to major topics in neuroscience at the secondary level. Teaching students about neuroscience exposes students to concepts such as neural communication, brain structure and function, and learning processes which are documented to have positive influences on students (MacNabb et al. 2006; Dubinsky 2010). Yet, the invisible biological processes inherent in neuroscience make neuroscience learning challenging. For example, communication between neurons, which is achieved through the process of neurotransmission at synapses, requires understanding of both electrical and chemical signaling. Students' inability to observe these complex processes at the molecular level contributes to the difficulty of understanding neuroscience content.

Given that classroom instruction that utilizes technology-based pedagogy can support student learning of advanced and complex science concepts (Li & Tsai 2013; National Research Council 2011), mobile device-based games have great potential to motivate students and improve neuroscience learning. Specifically, games or apps on hand-held devices can make use of visualizations and interactive simulations of neurobiological processes such as the communication between neurons within the human body. However, few apps or online resources are available to teach neuroscience (e.g., Chudler & Karikari 2015; Schotland & Littman 2012). Yet, research studies conducted in the context of these limited number of neuroscience games provided evidence supporting the use of games for neuroscience learning (Schotland & Littman 2012). Thus, traditional approaches to teach neuroscience may be rethought and new ways to integrate games into classroom instruction should be considered.

This study focuses on the use of an interactive mobile app for teaching neuroscience in secondary level courses. The purpose of this study was to compare the multiple ways a mobile device educational game (iNeuron), could be used to support the teaching of

neuroscience in high school classrooms in order to better understand what game features might make it an effective learning tool. The main research questions were the following:

1. How effective is iNeuron at improving student neuroscience content acquisition compared to standard instruction?
2. How is student learning affected by collaborative play in an educational game?
3. How is student learning affected by student-driven sequencing in an educational game?

Literature Review

There is a growing body of literature that suggests that mobile apps and games can be used to increase engagement and learning of science subjects (National Research Council 2011). Most mobile apps are designed as games since the interactive nature of games increases students' motivation to learn. Developers create classroom games for the specific purpose of supporting student learning rather than for entertainment purposes alone. These "serious" learning games "are designed to accurately model science or simulate scientific processes, and interactions within the virtual world of the game are governed by established scientific principles" (National Research Council 2011, p. 10). Within science education, a wide variety of serious games have been developed and tested at the middle school level (e.g., Anderson & Barnett 2013; Bresler & Bodzin 2016; Israel, Wang, & Marino 2015; Klisch et al. 2012; Schrader et al. 2016) and secondary school level (e.g., Annetta et al. 2009; Klisch et al. 2012; Sadler et al. 2013).

While science content and the learning goals are critical in the development of serious games, science education games should be grounded in theories and research on learning (Gee 2005; National Research Council 2011). Decades of research have shown that students learn science best when they actively engage in their learning (Dori & Belcher 2005; Freeman et al. 2014; National Research Council 2005, 2007). Within the context of science games, engaging students in the process of science learning means that students use digital technologies that support inquiry-based and active learning approaches. Such serious games allow students to not only increase their content knowledge but also allow them to practice scientific processes such as asking questions, carrying out investigations, and constructing explanations (Linn et al. 2010; *Next Generation Science Standards (NGSS)*, *NGSS Lead States* 2013). No deep science content learning takes place unless students engage in the key practices of science. In line with these research findings, three evidence-based pedagogical approaches guided app development in order to facilitate inquiry-based, active learning: *scaffolding*, *collaborative learning*, and *learner control*.

Within the context of STEM education, scaffolding is an approach to instruction in which students are guided in acquiring knowledge and solving inquiry-based problems that would otherwise be too difficult for them (Quintana et al. 2004). Scaffolds are often described as a unique way of providing appropriate guidance to students to reduce frustration and maximize learning (Pea 2004). Embedding learning scaffolds for successful challenge completion is an important design feature for serious games (National Research Council 2011). Scaffolds built into a game are supportive structures such as just-in-time hints,

prompts, or immediate feedback that allow learners to make progress when they encounter a problem. Scaffolds embedded in games enhance student scientific thinking and learning (Chang 2017; Hmelo & Day 1999; Nelson 2007; Rieber et al. 2004; White & Frederikson 1998). In addition, Molenaar et al. (2014) demonstrated that scaffolds in group play allow students to participate in social metacognitive interaction and thus support collaborative learning. Thus, each inquiry-based neural circuit challenge was preceded by a scaffolded lesson to introduce and reinforce new concepts and vocabulary. Students received hints and immediate feedback throughout the app, and each new challenge built upon the concepts learned in preceding challenges until students were able to construct relatively complex neural circuits on their own.

The application of collaborative strategies also supports meaningful science learning (Dillenbourg 1999; Gokhale 1995). Games that allow students to work collaboratively provide them opportunities to solve complex scientific problems in controlled situations (Ketelhut et al. 2006). Each student has a separate hand-held device but the simple data exchanges among students enable students to collect, share, or analyze scientific data. Participatory games motivate students and enhance learning (Roschelle 2003). In iNeuron, students could connect their devices and collaboratively build neural circuits. These same challenges could also be completed individually. This provided the opportunity to test the hypothesis that collaborative play in the context of a mobile device game would enhance student learning relative to individual play.

Game play also provides opportunities to test the effectiveness of learner control. Within the context of technology-based learning environments, learner control is defined as the ability of the learner to control the display of information or to have multiple ways to interact with information (Scheiter & Gerjets 2007). Different types of learner control can be distinguished, including control over sequencing information, selecting content, pacing of content, and timing and location of e-learning (Sorgenfrei & Smolnik 2016). The learner control afforded by technology-based learning environments is widely considered to be an advantage over standard forms of instruction, but empirical tests of its impact on learning have been mixed, suggesting a U-shaped dose-response curve (Sorgenfrei & Smolnik 2016). Reviews and meta-analyses have noted the difficulty of synthesizing the research on learner control given the wide variety of implementations from online employee training and MOOCs to other computer-based school learning (Dillon & Gabbard 1998; Karich et al. 2014; Niemiec et al. 1996; Sorgenfrei & Smolnik 2016). For instance, Karich et al. (2014) found that the overall effect size of learner control was near zero, but noted a moderate effect size ($g = 0.46$) for educational technologies that were comprehensive programs of instruction rather than review or extensive practice. Because iNeuron was designed to provide comprehensive basic neuroscience instruction, we chose to test the role that learner control in the form of sequencing might play in students' success in learning neuroscience concepts. Two versions of the app were compared; one with a fixed challenge order and one in which students controlled the challenge sequence within a content section.

Rigorous research on games is limited, and studies examining how games affect student learning vary in quality (Clark et al. 2015; National Research Council 2011; Vogel et al. 2006). Although studies have documented the effectiveness of games to support student

learning, game researchers often do not include a control group (e.g., Israel et al. 2015). In other studies, games were embedded in larger curriculum units, thus confounding the effects of curriculum materials and game (e.g., Sadler et al. 2013). Most studies use pre- and posttest data to measure student learning, but often neglect to perform analyses appropriate for designs in which randomization does not occur at the student level (Clark et al. 2016). In addition, typical game testing is unidimensional, not reflecting the many ways a game can be used in a classroom setting.

Taken together, design-based research approaches are needed in developing science games. Iterative cycles of design and research should contribute to the game development, allowing researchers to investigate the relationship between various aspects of game use and student learning with appropriate statistical analyses (Design-based Research Collective 2003). The current study employed a several-year, iterative approach including multiple rounds of classroom testing, student and teacher feedback and app redesign. This process culminated in the formal in-class testing reported here. While many studies have compared the effectiveness of specific instructional technologies against standard instruction, empirical research addressing how various game elements impact learning is needed. Our single-study design explores multiple implementations of the same app to address how inclusion of collaborative learning or learner control may contribute to the effectiveness of educational technology.

Methods

Game Context and Content Development

iNeuron is an iOS mobile app that teaches the basics of neuroscience through a combination of scaffolded lessons that reinforce content knowledge and inquiry-based challenges in which students build neural circuits to solve problems. The app includes both individual play and group play, in which students can connect their devices and solve challenges collaboratively. iNeuron was developed for high school-level science courses with neuroscience curriculum, such as biology, anatomy and physiology, and psychology.

Students navigate through the main map and complete a series of challenges in each section (Fig. 1). *Lesson* challenges provide students with scaffolded instruction on neuroscience topics, introducing new vocabulary and reinforcing concept knowledge through interactive exploration of screen content and through formative assessment multiple-choice and open-ended questions. *Circuit-building* challenges provide students with inquiry-based simulations, in which they must solve problems such as “*make this biceps muscle flex*” by creating circuits that connect neurons, the brain, muscles, and sense organs. Each section builds on the previous sections. Challenges increase in difficulty and complexity as students progress through the app. Students can choose to complete each *circuit-building* challenge individually or in a small group. In group play, up to five student devices can connect through Wi-Fi or Bluetooth, with different components of the challenge distributed among students in the group. Students then work together across their devices to collaboratively build each circuit and solve the challenge.

A systematic, iterative process was used to develop app content. The app was built as a standalone application using a peer-to-peer distributed application architecture for communication across devices. The internally developed Framework for Originating Responsive Game-based Education (FORGE) was used to construct app content. FORGE is a rapid-prototype, development platform which combines lesson plans, subject content, and game structure into a deployed app that supports individual lessons and group-based collaborative challenges. FORGE separates the details of the subject being taught from the lessons and challenges used to teach it, and from the interactions available on the device. Among other software development efficiencies, this disconnection allows instructional designers to more effectively scaffold lessons. The FORGE toolset enables quick iterations of prototypes so that the design and development could be informed by multiple review cycles with teachers during informal classroom testing.

After the initial design, four rounds of classroom piloting were conducted. During piloting, issues with the app design and content were identified and student survey feedback data were collected to inform useful revisions prior to the final evaluation study. Furthermore, piloting resulted in refining the study protocol and items on the multiple-choice quiz instrument developed to assess neuroscience concept knowledge. The independent pilot samples included a total of 333 8th-12th grade students in 14 classrooms across six teachers in five schools and five school districts. Class types included general biology, general psychology, and *Project Lead the Way* Medical Detectives courses. Teachers were recruited for piloting during local teacher conferences attended by the researchers.

In each round of piloting, students played with the app for one to two class periods. A researcher was present in each classroom to help students, to record student questions and where they appeared to have difficulty, and to observe the general level of student engagement. Informal anonymous surveys were collected from students at the end of each pilot to collect student feedback and suggestions for improvement. Informal verbal feedback was also collected from each teacher. Anonymous user data were logged by the app and later utilized to determine patterns of student use such as the number of challenges completed, the time to complete each challenge, and the responses to in-app multiple-choice questions. All feedback informed revisions to the content after each round of piloting.

Participants

The study sample included 399 9th-12th grade students (227 female) in 20 classrooms across 10 urban and suburban schools, and 11 teachers from the Midwest region of the U.S. Class types included general, AP, and honors biology; general and AP psychology; anatomy and physiology; an elective brain course; and special education science. Students' self-reported ethnicity was 57% White, 19% African American, 10% Asian American, 2% American Indian, 2% Native Hawaiian or Pacific Islander, and 10% more than one race. 12% of students self-reported as Hispanic or Latino.

Recruitment took place during a neuroscience professional development workshop for teachers led by the researchers during the summer prior to study initiation. Eleven 9th-12th grade teachers who completed the two-week workshop on inquiry-based instruction in neuroscience volunteered to participate. During the workshop, teachers received instruction

on inquiry-based pedagogy in neuroscience and were introduced to the app. Informed consent was obtained from all teachers, school principals, district administrators, and students prior to the commencement of data collection for the study in compliance with the researchers' approved IRB process.

Study Design & Procedure

To address the main research questions, a randomized block design study was implemented. Classrooms were pseudo-randomly assigned to one of five experimental conditions: *individual-linear*, *individual-nonlinear*, *group-linear*, *group-nonlinear*, or *control*. Assignment was pseudorandom. Data collection visits were scheduled four different times over a seven-month period in order to collect data from the 10 schools that participated in the study. For each round of school visits we pseudorandomly assigned each participating classroom to one of the five experimental conditions without replacement, so that teachers with more than one classroom in the study had each classroom assigned to a different condition. This was done both to minimize teacher effects within a condition and to achieve a data set with roughly equal *N*'s for each condition. Students in *group* classrooms participated in guided group play, while students in *individual* classrooms did not. Students in *linear* classrooms played a version in which all challenges must be completed in a specific order, while students in *nonlinear* classrooms played a version in which they were able to sequence the challenges within a particular section in any order. Finally, students in *control* classrooms learned basic neuroscience concepts via standard instructional methods the teacher chose to employ. These were self-reported to the investigators and included lessons that were taught to teachers during the neuroscience professional development workshop (e.g., building a model neuron from beads).

Data collection in each classroom spanned four consecutive class days. On day 1, students took a 10-item multiple-choice pre-quiz on neuroscience concept knowledge. On days 2 and 3, students played the app for the entire class period, or learned neuroscience concepts via standard instruction in *control* classrooms. On day 4, students took the same 10-item multiple-choice quiz as a posttest, followed by a brief demographic and usability survey.

The study was designed to ensure that iNeuron was realistically implemented as a teacher would choose to use it in class. Based on conversations with participating teachers, the same researcher led classes on days 2 and 3. The researcher paused the class two to three times per class period to provide a brief verbal summary of the concepts as they were encountered. She verbally quizzed students on key points and answered student questions. To prevent variability in implementation and alignment of teacher effects with condition, the same researcher led all classes in all conditions in which the app was used; teachers led the control classrooms.

For classrooms in *individual* conditions, students spent both class periods on days 2 and 3 working individually on an iPad. They were instructed by the researcher that they could work alone or alongside students near them, as appropriate to each classroom's own peer interaction culture. No instruction was given to students regarding the group play functions. For classrooms in *group* conditions, students were given the same instructions on day 2 as students in individual classrooms. On day 3 however, the researcher asked students to form

small groups of two to five and guided the class through a set of instructions for successfully connecting their devices and navigating collaborative group play challenges. The group-play challenges required students to work and reason together to connect pieces across devices. All students were given devices which employed software that kept them from exiting the app. Examination of log data after the study identified and subsequently eliminated less than five instances where app use was inconsistent with the assigned experimental condition.

Classroom observation data were collected from classes using the app to confirm consistency of implementation of the study protocol and to record general observations about classroom behaviors. Two-thirds of the classrooms were observed by the same evaluator, and one-third by two colleagues. Observers followed a 14-item observation protocol to guide their open ended descriptive field note taking. The items focused on the consistency of the researcher's delivery across classrooms; including instruction about the iPad, app, and neuroscience knowledge, direct vs discovery instruction, percent of time spent in whole class, small group, or individual instruction, approaches to helping students solve problems, and student usability or content issues.

Data Sources

Neuroscience concept knowledge assessment.—To assess student knowledge of neuroscience concepts, a 10-item multiple-choice pre- and posttest instrument was developed. Each item on the assessment had four response alternatives. The instrument was multidimensional, and covered the topics of neuronal anatomy and function, neurotransmitters, neural communication at synapses, firing rate, threshold, inhibition, and feedback. The items were developed by subject matter experts in neuroscience at the local university to ensure their validity and compared against educational standards for neuroscience instruction in the state and nationally (e.g., Next Generation Science Standards). Items on the instrument ranged from defining basic vocabulary (e.g. neurotransmitters and types of neurons) to higher-order-thinking items that required students to interpret graphs of neural firing rates and to infer the effects of damage to part of a neural circuit.

The instrument was refined based on student response feedback collected during the later rounds of classroom piloting, in which students were administered the pre-post assessment prior to and after using iNeuron. Informal item analysis of student performance was conducted after piloting to optimize response alternatives and the wording of each item. The instrument was distributed to all teachers participating in the study prior to data collection to receive feedback on the appropriateness of the items for their students and courses. The multidimensional nature of the instrument was confirmed via exploratory factor analysis using oblique rotation. Item difficulty values ranged between 0.26 and 0.85, and all items had non-zero and non-negative discrimination values.

Logged in-app data.—The app logged every student action, providing detail on where students navigated, what sections of the app were started and completed, what actions they performed on each screen, and what steps they took in *circuit-building* tasks. Logged data

were matched to individual student in-app and pre- and posttest code names to preserve student anonymity. These data were used in multiple exploratory analyses described below.

Classroom observation data.—Extensive field notes for each class period were reviewed by the researchers and used to note any differences in how class time was allocated in each of the four implementation conditions.

Usability survey.—Students completed the System Usability Scale (SUS), a 10-item questionnaire with response options for each item ranging from Strongly Agree to Strongly Disagree (Brooke, 2013). This high-reliability scale ($\alpha = .92$, Sauro, 2011) was developed to quickly evaluate the usability of hardware, software, mobile devices, websites, and applications. Survey data were used to determine if any significant usability issues arose that may have impeded learning of neuroscience concepts.

Data Analysis

Neuroscience concept knowledge assessment.—To examine the effect of experimental condition on posttest scores, a mixed effects model was fit in which the posttest scores were regressed on the condition and pretest scores, controlling for the nesting in the data (students nested within teacher). Specifically, the tested model was:

$$y_{ij} = \beta_1 + b_j + \beta_2 X_{ij} + \beta_3 C_{1ij} + \beta_4 C_{2ij} + \beta_5 C_{3ij} + \beta_6 C_{4ij} + e_{ij}$$

This model states that a student's posttest score (y_{ij}) was a function of their pretest score (X_{ij}), which condition they were in ($C_{1i} - C_{4i}$, dummy coded to indicate treatment conditions), a random variable (b_j) representing the deviation from the population mean of the mean posttest score for teacher j , and the residual random variable (e_{ij}). Teacher (rather than classroom) was the variable selected because the pseudorandom assignment of condition was done at the level of teacher.

This model was fit in R statistical software (R Core Team, 2017) using the “nlme” package (Pinheiro et al. 2017). The model was fit once using raw pre- and posttest scores, and again using adjusted pre- and posttest scores based on individual student progress through the app (see next section for details on calculating student progress and adjusted scores). Utilizing a linear mixed effects model that controls for the nesting of data is especially important in experimental designs where randomization occurs at the teacher or classroom level, but the unit of measure is the student (because students having the same teacher are unlikely to be independent); furthermore, it is underutilized in the educational intervention literature in which analyzing intervention study results using ANOVA or t -test is more common (Clark et al. 2016; Theobald & Freeman 2014).

Logged in-app data.—The number and type of challenges started and completed was logged internally by iNeuron for each student and used to determine individual student progress in the app. Each item on the pre-post assessment was matched with a corresponding challenge, in which the information needed to answer that item was contained. If a student did not reach a challenge corresponding to an item on the assessment, that item was dropped

from their pre- and posttest scores to create adjusted pre- and posttest scores. These adjusted scores were used in a second linear mixed effects analysis (described above) to determine if any observed differences in the raw posttest scores by experimental condition varied when student app progress was considered.

Student navigation data were used to calculate distance vectors to determine differences in the paths students chose in the *linear* and *nonlinear* experimental conditions imposed. Vectors were calculated using the Hamming distance. The Hamming distance vector measures the discrepancy in order between the base vector (intended order of challenges) and user vector (student initiated order), and only counts the first time a challenge is started (repeated challenges are disregarded, Hamming, 1950). The distance vector was normalized to the number of challenges that the user initiated. For example, comparing a string of ordered challenges '12345' against the user's string of challenges '14325' would generate a vector value of 2. Normalizing this vector to the number of challenges initiated (5) would produce a normalized vector value of 0.4. A normalized vector value of 0 denotes that the student initiated challenges exactly according to the intended challenge order, whereas a value of 1 indicates the student initiated challenges completely out of order. Vectors values were matched with student data and combined into *individual*, *group*, *linear*, and/or *nonlinear* combinations. Differences among group means were detected using a one-way ANOVA with a Tukey test for multiple comparisons.

Student errors in circuit-building challenges logged by iNeuron were used to investigate whether students demonstrated transfer of knowledge across different challenges. We used the National Research Council's (2000) definition of transfer as "the ability to extend what has been learned in one context to new contexts." Because the app's content was completely within the neuroscience domain, we measured *near* rather than *far* transfer. Two pairs of circuit-building challenges were designed to test similar neuroscience concepts but in different contexts. The number of errors made during circuit-building were summed for each student and compared between each pair of challenges to determine if the number of errors decreased from the first to the second. Only students that completed both paired exercises were included in the analysis. Differences were detected using a comparison of proportions statistic.

Usability survey.—Descriptive statistics and ANOVA were used to analyze usability survey data.

Results

Consistency and Usability

Observation and usability results are reported here as a way of establishing the grounds for the claims made about student learning. Qualitative field notes from 17 classroom observations documented the high consistency of the overall directions and neuroscience instruction in the classrooms, especially for the presentations on neurons and synapses, communication among neurons, firing rates, synaptic strengthening and pruning, and learning. The majority of the researcher's time was spent with individuals and small groups answering their questions about neuroscience, the iPad or the app. The majority of her

instruction was on neuroscience concepts. On average, 30% of instruction was to the whole class, 50% exploring and answering questions individually and 20% assisting small groups on group challenges. For the individual condition, students were engaged by themselves or with just one or two other students, consistent with the real world classroom setting.

The SUS was completed by all students who used iNeuron to ensure that app functionality was not an impediment to learning neuroscience or to completing the challenges. No significant usability issues were encountered. Excluding *control* and special education students, 252 students completed the System Usability Scale. Scores ranged from 15 to 100 with a mean of 69.04 and a standard deviation of 17.2. This corresponds to a score of “average” per SUS guidelines, which indicates that student learning of neuroscience concepts was not impeded by usability issues. Although the SUS was not originally designed for testing the usability of mobile-based technology, studies have shown that SUS scores for mobile applications in the range 67.7 – 87.4 are typical of apps considered to be top-10 across all app categories (Kortum & Sorber, 2015). Overall scores on the system usability scale did not differ significantly between conditions, which was tested using a one-way ANOVA: $F(3) = 0.529$, $p = .663$, indicating that the comparisons of student learning by experimental condition were not affected by any differences in perceived usability.

The observation notes also served as a triangulating check on the SUS as another indicator of the robustness, usability, and functionality of the app. The observation protocol addressed two usability issues; technical problems or the level of frustration with iPad or the app. Resoundingly across all 17 observations, observers noted no frustrations with the app or the platform. A small number of observed problems were fixed by simply restarting the challenge. Observers gauged student engagement as positive and continuous with both the app and the content during most classroom periods. In five observations, students engaged more with the technology and content on day two compared with day one.

Neuroscience Concept Knowledge Assessment

For student performance on the neuroscience concept knowledge assessment (Table 1) posttest scores were numerically higher than pretest scores in all five conditions. Of the 399 students who participated in the study, 344 matched pairs of pre- and posttests were obtained. The 23 pairs from students in two special education science (SPED) classes were removed from the main sample and analyzed separately, for a total N of 321 used in the subsequent analyses. Each item on the pre-post assessment was scored correct/incorrect for a total possible score of 10.

To assess the significance of the condition effect, a reduced model excluding the condition dummy variables was fit and compared to the full model (presented in the “Data Analysis” section) using a likelihood ratio test. This allowed the following null hypothesis to be tested:

$$H_0: \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1: \text{One of the above is not equal to 0.}$$

The results of this test were significant ($\chi^2 = 33.42$, $df = 4$, $p < .0001$), indicating there was a significant difference in posttest scores by condition (Table 2). All four experimental conditions had significantly higher posttest scores than the control group, controlling for pretest score and the teacher effect. The intraclass correlation for students nested within the same teacher was 0.255, which indicates that approximately 25% of variation in the observed posttest scores was accounted for by students sharing the same teacher. This variability probably reflects the breadth of student level factors such as demographics and how much neuroscience was covered prior to working with the app.

A post-hoc analysis was performed to compare pairwise student performance among the different conditions (with Bonferroni's adjustment, Table 3). The only differences found were between the control group and all four experimental conditions, where the control group had consistently lower posttest scores. Because the control group was used as a baseline in the above analyses, a paired samples t -test was used to compare performance between pre- and posttest for *control* students. The results were significant, indicating that the control group also showed improvement in neuroscience concept knowledge ($t(59) = 2.04$, $p = 0.046$, $d = 0.32$). Finally, differences between *individual* and *group* conditions and *linear* and *nonlinear* conditions were examined. The results of likelihood ratio tests indicated that there was no difference between *individual* and *group* ($\chi^2 = 0.564$, $df = 1$, $p = 0.453$) or between *linear* and *nonlinear* conditions ($\chi^2 = 0.078$, $df = 1$, $p = 0.779$).

A post-hoc, exploratory analysis, using a mixed effects model, was performed to assess whether posttest scores, controlling for pretest scores and the effect of a teacher, differed by a student's sex, their ethnicity, and whether they identified as Hispanic. For the ethnicity analyses, very few students identified as American Indian, Native Hawaiian, or Pacific Islander ($N = 7$), therefore, two different ethnicity analyses were performed. The first analysis dropped these students entirely and the second analysis consisted of recoding ethnicity as white or non-white. There were no significant differences associated with gender ($\chi^2 = .1131$, $p = .7367$); Hispanic status ($\chi^2 = .0022$, $p = .9629$); ethnicity with American Indian, Native Hawaiian, or Pacific Islander dropped ($\chi^2 = .7937$, $p = .2846$) or ethnicity treated as white/non-white ($\chi^2 = 1.2861$, $p = .2568$) controlling for pretest score and teacher.

The two special education science classes were assigned to the *individual linear* condition. Because both classes were assigned to the same experimental condition and the sample size was relatively small ($N = 23$ pairs), pre- to posttest gains were measured using a paired samples t -test. The results showed a significant increase in scores from pre to post: $t(22) = 2.41$, $p = 0.025$, $d = 0.49$.

Overall, these results address each of the three main research questions: 1. Students who used iNeuron learned neuroscience concepts better than students who learned via standard instruction; 2. The degree of student learning was not affected or enhanced by individual vs. group play; and 3. The degree of student learning was not affected or enhanced by either a linear game order or a nonlinear game order. Subsequent analyses presented below utilized individual students' logged in-app data to provide further insight into the pre- and posttest performance results obtained above, and to explore additional evidence of learning.

Analyses of Logged Student App Data

Analysis of student progress by condition.—To provide additional evidence for research questions 2 and 3, and further explore student performance on the pre-post concept knowledge assessment, individual student progress data were compared for each of the four app use conditions. While the analysis of the posttest scores presented above indicates that there was no difference in student performance based on utilizing the group play functionality, classroom observation data suggested that students in the two *group* conditions may not have progressed as far through iNeuron as students in the two *individual* conditions. This may have been due to the amount of class time required to arrange students into small groups, to guide them through the instructions for group play, and/or for students to communicate with each other to solve *circuit-building* challenges. In addition, individual students in all conditions varied in the amount of time they took to complete sections of the app. Because of this, some students did not complete sections of the app in which they would have learned concepts that were covered on the pre-post assessment.

To test this, logged challenge-completion data were used to determine app progress for each student and aligned their progress with the concept knowledge assessment items. Progress for each of the four experimental conditions was measured as the number of assessment items students were eligible to answer based on their challenge completions (Fig. 2). Comparisons between conditions using a one-way ANOVA showed that students in both *group* conditions made significantly less progress through the app than students in both *individual* conditions. Table 4 provides detailed progress data by condition, including the special education science classes.

Analysis of pre-post concept knowledge using adjusted scores based on app progress.—To provide additional evidence for all research questions and to account for individual student progress through iNeuron, adjusted pre- and posttest scores were calculated for students in each of the four conditions based on their progress. If logged app data indicated that a student did not complete a challenge, the item on the test corresponding to that challenge was omitted from their pre- and posttest scores. These adjusted scores were then subjected to the same mixed effects model analyses described above to determine if the differences in posttest performance by experimental condition revealed in the previous analysis might change once app progress was considered. The comparison between the full and reduced model was again significant ($\chi^2 = 26.35$, $df = 4$, $p < .0001$), indicating there was a significant difference in posttest scores by condition (Table 5). Again, all four experimental conditions had significantly higher posttest scores than the control group, controlling for pretest score and the teacher effect. The intraclass correlation for students nested within teacher was 0.205, indicating that approximately 20% of variation in the observed posttest scores was accounted for by students sharing the same teacher.

A post-hoc pairwise comparison on condition (with Bonferroni's adjustment, Table 6) found differences between control and three of the four experimental groups, where control had consistently lower posttest scores. However, the comparison between the *control* and *group-linear* conditions, while showing the same numerical trend as the full analysis, was no longer significant.

Finally, differences between *individual* and *group* conditions and *linear* and *nonlinear* conditions were examined. No differences were observed between *individual* and *group* conditions, although there was a numerical trend toward larger learning gains for the individual conditions than the group conditions ($\chi^2 = 3.51$, $df = 1$, $p = 0.061$). There was again no difference between *linear* and *nonlinear* ($\chi^2 = 0.641$, $df = 1$, $p = 0.423$).

Overall, the results of the analysis of adjusted scores are comparable to the analysis of the raw pre- and posttest scores, which indicated that students who used iNeuron outperformed students in the control group on the concept knowledge assessment. A notable exception is the non-significant post-hoc comparison between the *group-linear* condition and the *control* condition in the analysis of the adjusted pre- and posttest scores. One explanation for this result may be a lack of statistical power; as shown in Figure 2 and Table 4, the *group-linear* condition made significantly less progress than the other conditions and consequently had the fewest items included in students' adjusted pre- and posttest scores (6.19 items out of a possible 10 compared to 10 items out of 10 for the raw scores). The *group-linear* condition also had the smallest *N* of students compared to the other three conditions.

Transfer of key neuroscience concepts.—As further evidence for learning, logged in-app performance data were used to compare the number of errors students made when completing comparable circuit-building challenges (Fig. 3). For example, in the challenge *Flex Biceps* students were challenged to build a neuronal circuit to make the biceps muscle flex. In this early exercise, 56% of students made one or more errors. In contrast, only 29% of students made one or more error in the later challenge *Speak Up*, which challenged students to build a similar circuit to make muscles surrounding the mouth flex ($p < 0.001$, Figure 3). Similarly, a significant reduction in student errors was observed in comparable challenges *Feel the Sensation* and *A New Sensation*, from 80% to 33% respectively, which both challenged students to build sensory circuits ($p < 0.001$, Figure 3). These data document student knowledge transfer, as students made fewer errors while building circuits in later challenges that had similar neuroscience principles but were situated in a different context. This suggests that the scaffolded and interactive lessons contributed to students' understanding of neural circuit principles.

Linear vs. nonlinear activity.—To further address the effect of navigational route on neuroscience learning, log data were used to track the extent to which students in the *nonlinear* condition actually utilized their capability to navigate freely within each app section. Students in the *linear* condition were forced to execute challenges in a prescribed order, but still had the ability to revisit challenges already completed. Hamming distance vector values were obtained for each student, which measures the distance between an individual's navigation pattern and a predetermined linear navigation pattern. Students in *nonlinear* classrooms had greater than three times the Hamming distance compared to students in *linear* classrooms ($p < 0.001$, Figure 4). These results confirm that the experimental manipulation resulted in measurably different sequences for encountering content and completing challenges.

Discussion

The goal of this study was to examine student learning of neuroscience concepts using a mobile device game, and explore what game features might optimize student learning. The results of this study indicate that the app was an effective educational tool for secondary students in life science classrooms. After two class periods of use, students demonstrated higher posttest scores on a multiple-choice neuroscience assessment compared to students in a control group who received standard instruction, controlling for pretest scores. Student posttest performance was higher than control for all four conditions of app use, indicating that it can be implemented in the classroom flexibly to facilitate neuroscience instruction. One novel aspect of the experimental design was the comparison among four different implementations of the app, providing students with opportunities for peer-peer collaboration and sequencing control. Surprisingly, incorporation of these two evidence-based instructional approaches did not result in enhanced student performance relative to implementations without these features. The logged navigation data confirmed that students in *nonlinear* classrooms took advantage of the increased navigational freedom within the app. Choosing their own sequence did not diminish their learning. Navigation and task selection, two dimensions of learner control theory inherent in the sequencing choices made in the *nonlinear* conditions, have both previously been associated with positive cognitive or affective e-learning outcomes (Sorgenfrei & Smolnik 2016). Sequencing choices in the app were constrained within themes, providing some level of scaffolding, important in science education (Lin et al. 2012; Devolder et al. 2012). Such scaffolding may have minimized the learner control effect. Students did not become frustrated with the app as previously demonstrated in some instances of learner control (Hara & Kling 2000; Scheiter & Gerjets 2007). Rather students remained engaged, as confirmed by both app progress and classroom observations. The continued engagement may be attributable to the progressively more difficult, inquiry-based neural circuit-building challenges. This problem solving component may have contributed more than learner control to the outcomes. Further investigation should differentiate contributions from these multiple components of game design. Alternatively, learner control may be more appropriate for older learners or online procedural training rather than the cognitive knowledge and skills assessed here (Hannafin 1984; Carolan et al. 2014; Karich et al. 2014). These confounds prevented adequately testing the hypothesis that implementing learner control in a mobile device game facilitated learning. In conclusion, while the sequencing control available to students in the *nonlinear* condition did not result in learning gains, it also did not detract from their successful challenge completion or knowledge acquisition.

Another unexpected result of this study was that implementing collaborative play did not facilitate learning relative to conditions in which students only played individually. This result is consistent with a large meta-analysis of digital game research (Clark et al. 2016) in which it was found that games with collaborative group play did not produce significantly larger student learning gains than single-player games. One possible explanation is that the additional class time needed to get students into groups and to provide instruction on how to engage in group play led to less time on task for students. The analysis of logged in-app data is consistent with this explanation. Students in classes that engaged in group play made less

progress through the app overall than students in classes that did not engage in group play. By limiting game play to two class periods rather than allowing enough time for students to complete all challenges, the full effects of cooperative play may not have had time to materialize. This implies that teachers must weigh the established benefits of collaborative learning against the cost of class time to implement it when designing instructional activities for their students. Alternatively, the informal but less uniform peer interactions permitted during individual play may have provided sufficient collaboration that comparisons between these conditions were minimized.

Scaffolding may be the critical design element, acting as the elephant in the room, that dominated the results in this study. The app was heavily scaffolded, both for procedural hints and neuroscientific background content. A prior version of the circuit simulation at the heart of the app challenges was an online free-play game without any scaffolds (Dubinsky & Al-Ghalith 2012). Many teachers were hesitant to use the free-play simulation in their classes because they lacked sufficient domain knowledge and confidence to analyze circuit behaviors collaboratively with students. In free-play simulations, right and wrong are relative to the players' initial goals, making on-the-fly in-class evaluations context dependent. iNeuron was created to provide context, structure and less uncertainty to simulation challenges and to facilitate its use and adoption. The success of these scaffolds was revealed by the evidence of knowledge transfer. Students made fewer errors when building circuits that tapped previously covered concepts presented in a novel context. While not tested directly, the included scaffolds (content and hints) may have provided so much support that learner control and group interactions became less important contributors to successful challenge completion and knowledge gain.

Evidence for the flexibility of iNeuron came from analysis of student performance in special education classes. While the study was not specifically designed to examine whether this app would be an effective learning tool for these students, two classrooms of special education science were included after a special education science teacher in the professional development neuroscience workshop volunteered to participate in the study. Even with a relatively small sample of 23 students, significant learning gains from pre- to posttest were demonstrated. These results were encouraging in that they show how a complex topic such as neuroscience can be effectively taught with an app to secondary students who face obstacles in the learning process.

The *control* condition provided a fair and rigorous comparison for the experimental conditions. Teachers in the *control* group participated in the same two-week professional development workshop attended by the experimental teachers. Informal interviews conducted with teachers after data collection indicated that the lessons taught in *control* classrooms included inquiry- and active learning-based lessons that teachers experienced during the professional development. This provides some confidence that the higher performance of students who used the app was not simply due to the benefits of its personal interactive nature compared to other learning techniques. Furthermore, all teachers were given access to the neuroscience pre- and posttest prior to data collection. This prepared them for the content they were expected to cover and ensured that the assessment was not overly fitted to the app content.

Overall, this design-based research study used cycles of development, testing and revision to produce a robust game-like learning experience that augments classroom instruction for neuroscience. The scaffolded structure produced learning gains in a variety of classroom contexts across a range of student abilities. While versatile enough to be used in group contexts, this was not necessary to exceed the individual learning experience. Student navigational paths through the app could vary and still produce the expected learning.

Limitations

One limitation of the current study was that teachers were not the classroom facilitators, reducing its ecological validity. A researcher led the experimental classrooms to maximize the fidelity of implementation and increase the validity of the comparisons. Teachers were however familiar with the app from their prior neuroscience workshop, and they were present during the two class periods of use. Most teachers were active in helping to answer student questions and facilitate student use of the app, much as they would have done if the researcher was not present.

The relatively short period of instruction and time between assessments were dictated by the limited amount of time that teachers could devote to participating in the study. Most students did not complete all the challenges within the two classes, resulting in limited data for more advanced challenges. A primary concern of educators is what students can learn and retain over longer periods of time, and these results do not address the issue of long-term retention.

Future Directions

Many elements have been identified as crucial components for design of successful serious educational games. These include a sense of fun, challenge, reward, feedback, problem solving, conflict, self-determination, and autonomy in addition to learner control, collaboration and scaffolding (Prensky 2001; Martens et al. 2004; Chen et al. 2010). These design dimensions most likely interact when producing conditions for maintaining student interest, motivation and learning. Binary, or even dual as in the current study, comparisons fail to capture the multiple interactions among these design elements (Clark et al. 2016; Israel et al. 2015; National Research Council 2011). Multivariate techniques assessed the extent that student learning improved with app use, considering design elements and student population characteristics. What is needed is an independent comparative evaluation of a population of games along all of these dimensions to uncover the interactions and nonlinearity of dosages that combine to produce successful classroom tools. This proposed meta-analysis should employ even more powerful statistical techniques.

The richness and detail of logged in-app data provide many opportunities to explore subtle aspects of student performance on the simulations, not captured by knowledge assessment. Such an analyses will inform app revisions as well as identify specific neuroscience concepts students find difficult. Understanding how brains change with learning provides a unique context for both teachers and students to gain metacognitive knowledge about how people learn. When teachers and students receive instruction on the neuroscience of learning, teachers' pedagogy and students' school performance and attitudes towards learning

improve (Blackwell et al. 2007; Dubinsky 2010; Dubinsky et al. 2013). As these ideas drive neuroscience teaching to become more widespread, designing approaches to handle misconceptions will be critical for preventing the creation and spread of new neuromyths (Macdonald, Germine, Anderson, Christodoulou & McGrath 2017).

Acknowledgments:

This research was funded by National Institutes of Health R44MH096674 to MM and JMD. We would like to thank Dr. Nelson Soken, Kyle Nelson, Todd Carpenter, and Adam Gordon for their contributions to this project. We would like to thank all teachers who participated in the piloting and study, and especially Jeff Thompson for his extensive feedback.

References

- Anderson JL, & Barnett M (2013). Learning physics with digital game simulations in middle school science. *Journal of Science Education and Technology*, 22(6), 914–926.
- Annetta LA, Minogue J, Holmes SY, & Cheng MT (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, 53(1), 74–85.
- Blackwell LS, Trzesniewski KH, & Dweck CS (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246–263. [PubMed: 17328703]
- Bressler DM, & Bodzin AM (2016). Investigating Flow Experience and Scientific Practices During a Mobile Serious Educational Game. *Journal of Science Education and Technology*, 25(5), 795–805.
- Brooke J (2013). SUS: a retrospective. *Journal of Usability Studies*, 8(2), 29–40.
- Carolan TF, Hutchins SD, Wickens CD, & Cumming JM (2014) Costs and Benefits of More Learner Freedom: Meta-Analyses of Exploratory and Learner Control Training Methods. *Human Factors*. 56(5)999–1014. 10.1177/0018720813517710 [PubMed: 25141602]
- Chang H-Y (2017). How to augment the learning impact of computer simulations? The designs and effects of interactivity and scaffolding. *Interactive Learning Environments*, 25(8), 1083–1097.
- Chen K-C, Jang S-J, & Branch RM (2010). Autonomy, affiliation, and ability: Relative salience of factors that influence online learner motivation and learning outcomes. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 2(1), 30–50.
- Chudler EH, & Bergsman KC (2014). Explain the brain: websites to help scientists teach neuroscience to the general public. *CBE-Life Sciences Education*, 13(4), 577–583. [PubMed: 25452479]
- Chudler EH, & Karikari TK (2015). Neuroscience for kids. *Journal of Undergraduate Neuroscience Education*, 13(2), R14–R15.
- Clark DB, Tanner-Smith EE, & Killingsworth SS (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. [PubMed: 26937054]
- Design-based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8.
- Devolder A, van Braak J, & Tondeur J (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, 28(6), 557–573.
- Dillenbourg P (1999). What do you mean by collaborative learning. *Collaborative-learning: Cognitive and Computational Approaches*, 1, 1–15.
- Dillon A, & Gabbard R (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research*, 68(3), 322–349.
- Dori YJ, & Belcher J (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? *The Journal of the Learning Sciences*, 14(2), 243–279.

- Dubinsky JM (2010). Neuroscience education for prekindergarten–12 teachers. *Journal of Neuroscience*, 30(24), 8057–8060. [PubMed: 20554856]
- Dubinsky JM, & Al-Ghalith G (2012). Virtual Neurons. Retrieved on July 1, 2017 from brainu.org/lessons/virtual-neurons.
- Dubinsky JM, Roehrig G, & Varma S (2013). Infusing neuroscience into teacher professional development. *Educational Researcher*, 42(6), 317–329. [PubMed: 26139861]
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, & Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Gee J (2005). Good video games and good learning. *Phi Kappa Phi Forum*, 85(2), 33–37.
- Genetic Science Learning Center. (2013, 8 30) Mouse Party. Retrieved February 15, 2017, from <http://learn.genetics.utah.edu/content/addiction/mouse/>
- Gokhale AA (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, 7(1), Retrieved from: <http://scholar.lib.vt.edu/ejournals/JTE/v7n1/gokhale.jte-v7n1.html>.
- Hannafin MJ (1984). Guidelines for using locus of instructional control in the design of computer-assisted instruction. *Journal of Instructional Development*, 7(3), 6–10.
- Hara N & Kling R (2000) Students' distress with a webbased distance education course: an ethnographic study of participants' experiences. *Information, Communication, and Society* 3, 557–579.
- Hmelo C, & Day R (1999). Contextualized questioning to scaffold learning from simulations. *Computers & Education*, 32(2), 151–164.
- Israel M, Wang S, & Marino MT (2016). A multilevel analysis of diverse learners playing life science video games: Interactions between game content, learning disability status, reading proficiency, and gender. *Journal of Research in Science Teaching*, 53(2), 324–345.
- Karich AC, Burns MK, & Maki KE (2014). Updated meta-analysis of learner control within educational technology. *Review of Educational Research*, 84(3), 392–410.
- Klisch Y, Miller LM, Beier ME, & Wang S (2012). Teaching the biological consequences of alcohol abuse through an online game: impacts among secondary students. *CBE-Life Sciences Education*, 11(1), 94–102. [PubMed: 22383621]
- Klisch Y, Miller LM, Wang S, & Epstein J (2012). The impact of a science education game on students' learning and perception of inhalants as body pollutants. *Journal of Science Education and Technology*, 21(2), 295–303. [PubMed: 23926416]
- Ketelhut DJ, Dede C, Clarke J, & Nelson B (2006, 4). A multi-user virtual environment for building higher order inquiry skills in science. Paper presented at the 2006 AERA Annual Meeting, San Francisco, CA Retrieved from <http://muve.gse.harvard.edu/rivercityproject/documents/rivercitysympinql.pdf>
- Kortum PT, & Sorber M (2015). Measuring the Usability of Mobile Applications for Phones and Tablets. *International Journal of Human Computer Interaction*, 31, 518–529.
- Li MC, & Tsai CC (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, 22(6), 877–898.
- Lin TC, Hsu YS, Lin SS, Changlai ML, Yang KY, & Lai TL (2012). A review of empirical evidence on scaffolding for science education. *International Journal of Science and Mathematics Education*, 10(2), 437–455.
- Linn MC, Chang H-Y, Chiu J, Zhang H, & McElhaney K (2010). Can desirable difficulties overcome deception clarity in scientific visualizations? In Benjamin AS (Ed.) *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork*. New York: Routledge
- Macdonald K, Germine L, Anderson A, Christodoulou J, & McGrath LM (2017) Dispelling the Myth: Training in Education or Neuroscience Decreases but Does Not Eliminate Beliefs in Neuromyths. *Front. Psychol* 8:1314. [PubMed: 28848461]
- MacNabb C, Schmitt L, Michlin M, Harris I, Thomas L, Chittendon D, Ebner TJ, & Dubinsky JM (2006). Neuroscience in middle schools: a professional development and resource program that models inquiry-based strategies and engages teachers in classroom implementation. *CBE-Life Sciences Education*, 5(2), 144–157. [PubMed: 17012205]

- Martens R, Gulikers J, & Bastiaens T (2004). The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted Learning*, 20(5), 368–376.
- Molenaar I, Slegers P, van Boxtel C (2014). Metacognitive scaffolding during collaborative learning: a promising combination. *Metacognition Learning*, 9, 309–332.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: The National Academies Press 10.17226/9853.
- National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom*. Washington, DC: The National Academies Press.
- National Research Council. (2007). *Taking science to school: Teaching and learning science in grades K-8*. DC: The National Academies Press.
- National Research Council. (2011). *Learning science through computer games and simulations*. National Academies Press.
- Nelson B (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology*, 16(1), 83–97.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.
- Niemiec RP, Sikorski C, & Walberg HJ (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research*, 15(2), 157–174.
- Pea R (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, 13(3), 423–451.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, and R Core Team (2017). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–131, <https://CRAN.R-project.org/package=nlme>.
- Prensky M (2001). *Digital game-based learning*. New York: McGraw-Hill.
- Quintana C, Reiser BJ, Davis EA, Krajcik J, Fretz E, Duncan RG, Kyza E, Edelson D, Soloway E (2004) A Scaffolding Design Framework for Software to Support Science Inquiry. *Journal of the Learning Sciences*, 13(3), 337–386.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL <https://www.R-project.org/>.
- Rieber LP, Tzeng S, and Tribble K (2004). Discovery learning, representation, and explanation within a computer-based simulation. *Computers and Education*, 27(1), 45–58.
- Roschelle J (2003). Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning*, 19(3), 260–272.
- Sadler TD, Romine WL, Stuart PE, & Merle-Johnson D (2013). Game-based curricula in biology classes: Differential effects among varying academic levels. *Journal of Research in Science Teaching*, 50(4), 479–499.
- Sauro J (2011). SUSified? Little-Known System Usability Scale Facts. *User Experience Magazine*, 10(3). Retrieved from <http://uxpamagazine.org/sustified/>.
- Scheiter K, & Gerjets P (2007). Learner control in hypermedia environments. *Educational Psychology Review*, 19(3), 285–307.
- Schotland M & Littman K (2012). Using a computer game to teach young children about their brains. *Games for Health Journal*, 1(6), 442–448. [PubMed: 24761319]
- Schrader PG, Deniz H, & Keilty J (2016). Breaking SPORE: Building Instructional Value in Science Education using a Commercial, Off-the Shelf Game. *Journal of Learning and Teaching in Digital Age (JOLTIDA)*, 1(1), 63–73.
- Sorgenfrei C & Smolnik S (2016) The Effectiveness of E-Learning Systems: A Review of the Empirical Literature on Learner Control. *Decision Sciences Journal of Innovative Education*. 14(2), 154–184.
- Theobald R, & Freeman S (2014). Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE-Life Sciences Education*, 13(1), 41–48. [PubMed: 24591502]
- Vogel JJ, Vogel DS, Cannon-Bowers J, Bowers CA, Muse K, and Wright M (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243.

White B & Frederiksen J (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3–118.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1. iNeuron organizational structure. A) Main navigational page. Students progress through seven sections, each covering a particular topic in secondary neuroscience curriculum. B) One of the section maps, containing several challenges that students must complete before moving on to the next section. C) Example screen from a lesson challenge. D) Example screen from a circuit-building challenge.

Fig. 2.

Progress through iNeuron differed by experimental condition, with students in group conditions making less progress over two class periods than students in individual conditions. The y-axis shows the number of assessment items (out of 10) that students were eligible to answer based on what challenges they completed. Students were included in this analysis if they had both completed pre- and posttest scores and had a user name that could be matched to their pre-post test data. Differences were detected using a one-way ANOVA with a Tukey test for multiple comparisons; * $p < 0.05$, *** $p < 0.001$

Fig. 3.

Error Analysis of similar challenges. A. Pairwise comparison of the number of students who made one or more errors in circuit building exercises Flex Biceps and Speak Up (N = 263). B. Pairwise comparison of the number of students who made one or more errors in circuit building exercises Feel the Sensation and A New Sensation (N = 132). Only students who completed both challenges were included in this analysis. Challenges in black preceded sections in grey. Differences were detected using a comparison of proportions statistic; *** $p < 0.001$

Fig. 4.

Box and whisker plots of distance vectors of navigational routes through iNeuron across experimental conditions. Boxes represent the interquartile range (25–75%). Lines represent first and fourth quartile. Group means are represented by +. Differences among means were detected using a one-way ANOVA with a Tukey test to correct for multiple comparisons;

*** $p < 0.001$

Table 1

Pre- and Posttest Scores for Neuroscience Concept Knowledge Assessment

Condition	N of pairs		Pretest		Posttest	
	<i>N</i>	%	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Control</i>	59	18	4.49	1.61	5.05	1.90
<i>Individual linear</i>	68	21	4.29	1.63	6.18	2.29
<i>Individual nonlinear</i>	68	21	4.43	1.75	6.53	2.18
<i>Group linear</i>	56	17	4.04	1.65	5.29	2.15
<i>Group nonlinear</i>	70	22	4.40	1.65	5.66	2.17
<i>SPED science class</i>	23	100	3.04	1.55	3.87	1.82

Table 2

Estimated Parameters for The Full Model

	Estimate	SE	df	t-value	p-value
Intercept	2.59	0.55	306	4.74	< .001
Pretest Score	0.37	0.06	306	5.88	< .001
Individual linear	1.97	0.39	306	5.03	< .001
Individual nonlinear	2.64	0.51	306	5.18	< .001
Group linear	2.4	0.53	306	4.49	< .001
Group nonlinear	2.01	0.43	306	4.67	< .001
Teacher Variance (σ_b^2)	1.07				
Residual Variance (σ_e^2)	3.13				

Table 3

Post-Hoc Analysis for Condition

	Estimate	SE	df	t-value	p-value
Control - Individual linear	-1.97	0.39	306	-5.03	<.0001
Control - Individual nonlinear	-2.64	0.51	306	-5.18	<.0001
Control - Group linear	-2.40	0.53	306	-4.49	0.0001
Control- Group nonlinear	-2.01	0.43	306	-4.67	<.0001
Individual linear - Individual nonlinear	-0.67	0.45	306	-1.48	1
Individual linear - Group linear	-0.43	0.46	306	-0.93	1
Individual linear - Group nonlinear	-0.05	0.38	306	-0.12	1
Individual nonlinear - Group linear	0.24	0.46	306	0.53	1
Individual nonlinear - Group nonlinear	0.62	0.36	306	1.74	0.828
Group linear - Group nonlinear	0.38	0.42	306	0.91	1

Table 4

Assessment Item Eligibility Based on In-App Progress

Condition	Mean Assessment Items Eligible to Answer	Standard Deviation	<i>N</i>
Group Linear	6.19	2.62	54
Group Nonlinear	6.60	2.98	67
Individual Linear	7.87	2.68	62
Indv. Lin. (SPED)	7.55	2.79	22
Individual Nonlinear	8.48	2.58	63

Table 5

Adjusted Pre-Post Scores: Estimated Parameters for The Full Model

	Estimate	SE	df	t-value	p-value
Intercept	0.94	0.51	306	1.86	0.06
Pretest Score	0.79	0.06	306	13.81	< .001
Individual linear	1.80	0.39	306	4.55	< .001
Individual nonlinear	2.09	0.50	306	4.16	< .001
Group linear	1.15	0.53	306	2.18	0.03
Group nonlinear	1.53	0.43	306	3.55	< .001
Teacher Variance (σ_b^2)	0.826667				
Residual Variance (σ_e^2)	3.209834				

Table 6

Adjusted Pre-Post Scores: Post-Hoc Analysis for Condition

	Estimate	SE	df	t-value	p-value
Control - Individual linear	-1.80	0.39	306	-4.55	0.0001
Control - Individual nonlinear	-2.09	0.50	306	-4.16	0.0004
Control - Group linear	-1.15	0.53	306	-2.18	0.3012
Control- Group nonlinear	-1.53	0.43	306	-3.55	0.0044
Individual linear - Individual nonlinear	-0.29	0.45	306	-0.64	1
Individual linear - Group linear	0.64	0.46	306	1.40	1
Individual linear - Group nonlinear	0.26	0.38	306	0.70	1
Individual nonlinear - Group linear	0.93	0.45	306	2.05	0.4117
Individual nonlinear - Group nonlinear	0.55	0.36	306	1.53	1
Group linear - Group nonlinear	-0.38	0.42	306	-0.90	1