Improving Achievement Using Digital Pedagogy: Impact of a Research Practice Partnership in New Zealand

Rebecca Jesson, Stuart McNaughton, Aaron Wilson, Tong Zhu & Victoria Cockle


To link to this article: https://doi.org/10.1080/15391523.2018.1436012

Published online: 06 Mar 2018.

Submit your article to this journal

Article views: 373

View Crossmark data
Improving Achievement Using Digital Pedagogy: Impact of a Research Practice Partnership in New Zealand

Rebecca Jesson, Stuart McNaughton, Aaron Wilson, Tong Zhu, & Victoria Cockle
University of Auckland

Abstract

This present study reports the impact of a research–practice partnership among nine collaborating schools and researchers in Auckland, New Zealand. The goals of the partnership were to refine digital instruction in ways that would result in improved learning processes and achievement. The partners employed a design-based research approach to iteratively develop and test changes to the pedagogy. Results indicate moderate to large effect sizes in writing, and small effect sizes in reading and mathematics. Results also indicate increasing effects over time in reading and writing, but not in mathematics. Educational effects included greater use of discussions by teachers and greater use of open-ended activity types. These outcomes are discussed in relation to features of the partnership as well as digital pedagogy. (Keywords: design-based research, 1:1 device use, school improvement, research practice partnership)

Digital technologies are an integral part of day-to-day lives across the world, and are becoming common in schools. In New Zealand, where schools are self-managing, schools are increasingly seeking to offer ubiquitous access to digital devices for learning. Such one-device-per-child provision enables use of a digital device whenever it is required, and digital use that is positioned as a background norm, rather than a specific innovation (Berger-Tikochinski, Zion, & Spektor-Levy, 2016).

This study is situated in schools serving a low socioeconomic community in Auckland. New Zealand does not have school districts; however, school collaboration occurs voluntarily. The public schools in this area have “clustered” together and formed an educational trust, which supports families in buying low-cost digital devices for their children; it administers service agreements with infrastructure providers, and resources professional development for teachers. Like similar initiatives, the goals are to support equity of access to technology for learning, enhanced academic outcomes, and preparation for future careers (Penuel, 2006). Pedagogical shift is driven through a shared digital pedagogy, implemented across all the schools, in which student engagement in learning is valued. The pedagogy is conceived as a learning cycle, in which students build curriculum knowledge, often through reading, discussion, or research, based on which they then create a digital artifact. Such artifacts are often multimodal, and serve to synthesize learned content into an individual creation. Students then post their digital creations to an individual blog, thereby sharing their learning with their family or the wider public, who are encouraged to view and comment on these posts. The teachers in the cluster have each created a class site, where students access their tasks, texts, and resources, as well as timetables and any other information they may need. In this way, both learning and teaching are intended to be visible to the community. Schools also host training sessions for parents, so that they learn how to access and comment on their children’s blogs. The cluster of schools has partnered with the authors in a research–practice partnership. The role of researchers has been to collect and analyze data about the effects of the program on teaching and learning and to engage in collective sense-making with our practice partners.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujrt
It is recognized that digital devices do not necessarily improve learning. Achievement outcomes from most large-scale digital implementation projects report modest or mixed results (Fleischer, 2011), with variability within and across schools in implementation and in pedagogy (e.g., Lowther, Inan, Ross, & Strahl, 2012). Zheng, Warschauer, Lin, and Chang’s (2016) recent meta-analysis suggests an overall small effect on achievement \( (d = 0.16) \), with effects for writing \( (d = 0.20) \), English language arts \( (d = 0.15) \), and mathematics \( (d = 0.17) \), and nonsignificant small effect sizes for reading \( (d = 0.12) \). Thus, while digital technologies are becoming the background tools of everyday experience, the search continues for ways to implement these in schools such that students’ learning is enhanced in line with program goals.

**Theoretical Perspective**

In considering how digital tools might enhance learning, we draw on sociocultural concepts (Rogoff, 2003; Goncu & Gauvain, 2012). From this perspective, development is a mutually constitutive interaction between learners and more expert people, in which learners come to know the goals, actions, and conditions relevant to specific activities that define the ongoing practices within and across settings and groups (Rogoff, 2003). Participation within these activities reflects the nature of the resources and the mediation afforded by the context and specific tools within those settings. Learners are conceived as active: negotiating, internalizing, or appropriating the meanings available to them, thereby becoming increasingly expert (Goncu & Gauvain, 2012). Given this conception, two classroom processes are key to considering educational change. One is a need to understand new configurations of, resources for, and the functions of the tools associated with activities. The other is the need to understand altered interaction patterns associated with new features of that environment.

**Improving Learning in Digitally Enabled Learning Environments**

We have identified five key ways that activities in digital learning environments, supported by resources and tools, can provide improved opportunities for students’ learning. These opportunities relate to (a) efficiencies that maximize learning time, (b) affective learning including engagement, (c) a broader and higher level set of skills, (d) the use of productive as well as receptive learning processes, and (e) changed patterns of instructional interaction.

Ubiquitous digital access has been associated with opportunities for increased learning-focused time, with a corresponding increase in students’ production. Studies identify increased participation in composing processes (Lowther et al., 2012; Warschauer, 2009), and, in accordance with this, Zheng et al. (2016) reported that writing, editing, and gathering information from the Internet were the most common uses of laptops in the studies they reviewed. Arguably, such synthesis of multiple sources might be considered to be both a traditional task that can be achieved more efficiently digitally, and an activity that is increasingly required by a digital environment.

Another recurrent finding is that digital environments provide opportunities for increased engagement, widely defined, but including interest and enjoyment (Bebell & O’Dwyer, 2010; Suhr, Hernandez, Warschauer, & Grimes, 2010). Triangulating teacher interviews, student surveys, and classroom observations in the Berkshire Wireless Learning Initiative, Bebell and Kay (2010) reported considerable improvements in engagement (including participation, enthusiasm, motivation and on-task behavior) across 3 years. This often-reported impact on engagement is germane to efforts to improve provision in New Zealand schools that serve low-income communities, where disengagement from schooling is identified as a barrier to success, and increasing engagement in learning is identified as a necessary precondition for improving educational outcomes (Gibbs & Poskitt, 2010).

Skills related to, but usually not directly measured by, standardized tests are also identified as enhanced through digital learning (Warschauer, 2008). The identified skills include cognitive skills such as critical literacy or critical thinking skills, and associated independent learning skills such as personal organization, study skills, and study habits (Bebell & Kay, 2010), and information
skills such as organizing information, distinguishing between sources, evaluating information, argument writing, and presentation of findings (Spektor-Levy & Granot-Gilat, 2012). The broader set of skills required for interpreting digital sources seems to align with the picture of students independently gathering information from the Internet to compose or respond, as reported by Zheng et al. (2016).

The use of the digital environment to allow students to become producers as well as consumers of knowledge (McLoughlin & Lee, 2008) can also be identified in the nature of students’ learning activities. Investigations into what students compose using digital tools indicate the potential for increased leverage from productive activities as an approach to demonstrating one’s thinking (Mouza, 2008). Again, however, the potential is dependent on pedagogy. In their analysis of 154 digital artifacts in mathematics and science, Rodriguez, Frey, Dawson, Liu, and Ritzhaupt (2012) noted improved use of technology but a continuance of lower level tasks for learning content. The authors argue that activity design should seek to extend students’ higher level skills, through challenging cognitive work using digital tools.

Digital implementation is also thought to have positive potential impact on instructional interactions (Grimes & Warschauer, 2008; Lowther et al., 2012). The authors of the Becta 5-year study in 28 schools reported improved quality of interactions, leading to more learner-initiated choices, and greater teacher responsiveness to needs (Somekh et al., 2007). In writing, Warschauer also reported shifts toward more purposeful, authentic tasks, as well as increased autonomy and collaboration (Warschauer, 2008; 2009).

Developing More Effective Digital Pedagogy Across Multiple Sites
While small-scale studies have been able to document changes in learning activities and interactional properties, these possibilities do not necessarily translate into effects in larger rollouts, with the associated variability in implementation (Bebell & Kay, 2010). Accordingly, a number of scholars have called for studies to focus on what counts as effective pedagogy within digitally enabled environments, and how teachers might be supported to design instruction thatcapitalizes on potential advantages while avoiding pitfalls (Bebell & O’Dwyer, 2010; Dunleavy, Dexter & Heinecke, 2007). A generalized collective approach to redesigning a curriculum is exemplified in the Becta initiative, where the learning gains sit within an explicit pedagogical framework designed for ubiquitous digital contexts (Underwood, 2009). In the present study, a shared digital pedagogy was used by teachers across the cluster, as they sought to use ubiquitous access to redesign their instruction to best promote learning. The pedagogy was conceived as enhancing engagement through student creation of digital artifacts about learning, which were posted on students’ blogs for their family, peer, or public audiences. This visibility of learning and teaching was intended to enhance the partnership between families and school, as well as to provide motivation for students who can respond to comments as well as see the numbers of views on their blogs. A research-practice partnership, where researchers, school leaders and teachers jointly worked to interrogate the research data to continually improve instructional designs supported this endeavor.

The Research–Practice Partnership
In this study, we used a research–practice partnership (RPP) (Snow, 2015) approach to undertake design-based research (Anderson & Shattuck, 2012) to jointly address an agreed problem. Snow identifies four principles that are hallmarks of a RPP. The first principle is that knowledge comes from both research and practice. Valuing these two sources is designed to increase the amount and quality of knowledge, while acknowledging both partners’ commitments and constraints. The partnership in this study was framed around a design-based approach that has previously contributed to the raising of literacy achievement in New Zealand, the Learning Schools Model (McNaughton, Lai, Jesson & Wilson, 2013). Whereas previous iterations of the model relied on researchers to design and refine interventions, in this iteration, schools had begun to intervene independently, and
the partnership was employed to refine and redesign aspects of that intervention to drive continued improvement.

A second characteristic of an RPP approach is that it emerges from the pressing concerns of practitioners. In this case, the schools were concerned to address an achievement issue by promoting increased engagement through the introduction of digital pedagogy. The partnership with researchers was instigated to contribute to solving achievement issues. While improvement efforts by the schools were focused particularly on improving digital pedagogy in writing, the research team also tracked student achievement data in other tested areas (reading and mathematics).

A third principle of RPPs is that improving practice requires evaluating the effectiveness of innovations. Our approach to evaluation was premised on the understanding that contextual variation would be a key lever for improvement, and that effective instructional practices needed to be developed from evidence within the context, combined with a strong theoretical rationale. The potential benefits of this type of partnership are those of buy-in, relevance, and utility. It is intended that teachers will better understand the research and its implications because they helped develop it; that the research will be helpful for teachers because it is grounded in their own schools; and that findings will be more credible to other teachers because the studies are done with students in their own schools (Coburn, Penuel, & Geil, 2013).

Snow’s fourth principle is attention to systemic change. Cluster-wide and school-level changes were driven in this study through the formative use of the collected data by the participants in professional learning communities (PLCs) within and across schools, based on which teachers fine-tuned their practices and analyzed the effects of these refinements to ensure that changes constituted an improvement. The design-based approach allowed ongoing refinements based on shared hypotheses discussed in PLCs to occur iteratively throughout the partnership project.

In summary, our approach to partnering was to work with schools to solve a local problem (in this case achievement outcomes, particularly in writing), and through that partnership, to contribute to the understandings of the design of effective digital pedagogy. In keeping with principles of design-based research (Anderson & Shattuck, 2012), we were committed to impacting achievement through the digital design, but also to advancing knowledge through the research process. Here we report the outcomes of the study as trajectories of change over time, to help inform our understanding of more effective digital instructional designs.

Methods

Context
The project was designed as a partnership involving schools in a self-formed “cluster” of nine schools in a low-income suburb of Auckland, New Zealand. The school sizes ranged from 95 to 632 students. Eight of the schools were primary schools (Years 1–8) and one was a high school (Years 9–13). The schools had worked together previously to attempt to raise achievement in the local area. One primary school had trialed student device ownership, judging it to be successful, based on learning gains in trial classes. In 2011, based on this initial trial, five other primary schools and the high school adopted digital learning for all their students in Years 4–13, moving from traditional approaches to one-to-one provision for students along with a shared pedagogy for teachers. In 2012, three more primary schools joined the group. Teachers developed Google Sites to provide activities, resources, texts, timetables, and any other information required by students to access learning. Students created written, visual, or multimodal artifacts to explain their thinking and posted these to individual blogs.

Participants
The schools serve a transient population of students (approximately 30% turnover each year). Students in the cluster at any given time were from a variety of ethnicities, including Pacific Nations
heritages (65%), New Zealand Māori heritage (25%), New Zealand European heritages (3%), and others, including Asian heritages (7%).

Research Design
Within the overall RPP, the research approach mirrored quasi-experimental intervention designs. Here we detail the outcomes of the first 3 years of the initiative, which, through partnership with researchers, iteratively refined the approach to improving educational outcomes for students. Student achievement data were collected twice yearly to monitor achievement throughout the study and to provide repeated measures. The initial (2012, Time 1) achievement levels for each cohort provided a cross-sectional baseline for systematic comparisons using the repeated measures. This enabled us to examine ongoing effects on achievement for longitudinal cohorts of students who were present for the whole intervention. Hierarchical linear models (HLMs) were used to examine the mixed effects over time for all students involved in the intervention to varying degrees, while accounting for variation in students’ initial levels, class and school.

Partnership Procedures
The partnership approach followed three broad phases: an initial profiling phase (for discerning patterns of instruction), an implementation phase (for refinement of instruction), and a sustainability phase (to embed change processes). Data analyses were interpreted by both partners through professional learning community (PLC) structures within the cluster. Data collection and oral feedback occurred twice each year: at the beginning of the year as a whole cluster staff meeting, and in Term 3 as a hui (large gathering) of key school members. Written analyses occurred twice yearly for the cluster and each school, and were synthesized yearly as a written report. Each feedback artifact (presentation or report) was discussed or contested in PLCs throughout the cluster, and action plans were refined at principal, cluster curriculum leaders, and staff and teaching team levels. The cycles were formative in that they contributed to further refinements to increase the effectiveness of the innovation.

Phase One: Developing a Profile and Shared Hypotheses
To contribute to understanding patterns of achievement generally, in the initial profiling phase, a range of data sources were collected as the basis for hypotheses which would drive instructional improvement. Patterns of achievement, aspects of classroom digital pedagogy, responses to leader questionnaires, and debrief interviews with teachers and students were analyzed to form a “profile” and are reported elsewhere (Jesson, McNaughton & Wilson, 2015). The evidence was used to develop understandings about how the nascent intervention was impacting outcomes, whether there were identifiable areas of high impact, and what the features of more effective teaching might be. At the end of the year, hypotheses for improvement were posited and shared with the schools. The hypotheses were drawn from literature that indicated possibilities for enhanced learning provided in the digital environment, combined with the profile evidence, which indicated how these possibilities were already being used in the schools. The (summarized) hypotheses were that outcomes would be enhanced if schools and teachers:

- Capitalized on their shared vision by further developing a fully shared focus on accelerating student achievement.
- Developed a theory of action that translated digital affordances into increased outcomes for students by growing effective practices.
- Exploited the high levels of student engagement by increasing the cognitive complexity of activities.
- Capitalized on affordances in efficiency by shifting to more interactions extending deep thinking.
The profiling data and the resultant hypotheses were discussed in detail with principals, leaders, and teachers in feedback sessions and in PLC follow-up sessions. Each group “unpacked” the hypotheses, interrogated their validity, and collectively created work plans and goals for how instruction might be refined to achieve pedagogical shift based on the hypotheses. In this way, practice knowledge formed the basis for the planned refinements to pedagogy in line with research hypotheses (see Table 1).

**Phase Two: Refinements to Instruction and Analysis of Variability**

In the second year (2013), cluster leaders continued to use existing PLC structures with teachers to translate the shared hypotheses into actions. The research team monitored the implementation through classroom observations and continued analysis of student achievement data. Formative feedback to the cluster included evidence of levels of implementation of digital pedagogy, numbers of classes with accelerated achievement, and shifts in patterns of classroom instruction consistent (or inconsistent) with Hypotheses 3 and 4.

The iterative cycles of data collection established in the profiling phase were continued. Researchers’ analyses of student achievement data contributed to the schools’ ability to effect Hypothesis 1. In addition, to contribute research into Hypothesis 2, effective teachers were invited to be “case studies” for in-depth analysis. Based on the schools’ focus on writing, and on the achievement data that continued to highlight writing as an underperforming area, effective teachers (those with greatest average student gains in writing) were selected as case studies (Jesson, McNaughton, Rosedale, Zhu, & Cockle, 2018). Analyses of case-study teachers’ Google sites and students’ blogs were conducted in an effort to explain why these teachers might get better than average student progress results. Consistent with the features of a RPP, this aspect of the research was intended to use the variability inherent in the context to drive further improvement, specifically in the teaching of writing.

**Phase Three: Iterative Refinement Through Sustainable Processes**

During the final phase, themes and examples arising from the case studies were added to the analyses presented to school and cluster leaders. Again, through PLCs, the practices of effective teachers of writing were integrated into the discussion of refinements to practice. By this time, processes for the analysis and translation of research findings into actions for teachers were embedded by the cluster and attention turned to the maximizing the impact of these structures. In particular, cluster leaders invested efforts, through cluster-wide planning linked to school planning, so that professional learning within the cluster contributed to teachers’ ability to work toward the shared hypotheses. Teachers and leaders within the school used PLC structures to share the results of their own inquiries into practice, and to align these with the shared hypotheses. This feature of the design aligned with the RPP concern to work coherently at multiple levels. During this time, attention also turned to emerging practice problems to solve: the continued brake on writing achievement posed by the summer break, and patterns of a plateau at higher year levels in reading and mathematics. These ongoing issues for the cluster formed the bases of subsequent partnerships.

**Measures**

**Assessment Tools for Teaching and Learning (asTTle)**

The e-asTTle suite of tools (https://e-asttle.tki.org.nz) was selected by the schools as a shared standardized measure. The tool draws from banks of items aligned to the New Zealand Curriculum. Items are scored using single-parameter Item Response Theory (IRT) and transformed to a scale score. Results are presented relative to curriculum outcomes and national norms of performance. Using the tool, common tests for reading, writing, and mathematics were developed, with parallel tests for the beginning and end of each year. Tests were designed with similar construct composition and item difficulty ranges across the study and were administered in the same week in each school, with between- and within-school moderation processes included in each testing cycle.
<table>
<thead>
<tr>
<th>Group</th>
<th>Participants</th>
<th>Purpose</th>
<th>Activities</th>
<th>Artifacts</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole cluster feedback</td>
<td>All staff from all schools</td>
<td>Provide research feedback about cluster as a whole to all teachers and leaders</td>
<td>Presentation and discussion of student achievement and teaching data patterns at cluster level up to end of previous school year. Discussions for evaluation and reflection, hypothesis generation and preliminary goal setting</td>
<td>Data presentation</td>
<td>Yearly (February)</td>
</tr>
<tr>
<td>Principals’ PLC</td>
<td>Principals of all schools</td>
<td>Principal learning, cluster and school coherence, strategic direction</td>
<td>Data discussion, principal knowledge sharing, cluster planning</td>
<td>School and cluster student achievement data, classroom observation analyses cluster plans, research presentations, milestone reports</td>
<td>Two full PLC days in addition to four regular meetings per year</td>
</tr>
<tr>
<td>School leaders’ PLC</td>
<td>Deputy and associate principals or curriculum leaders (including year-level and/or learning-area leaders)</td>
<td>Leader learning, dissemination to teachers, cross-school collaboration, feedback from leaders about activities in their schools</td>
<td>Data discussions, professional learning, presentations by school leaders, researcher presentations (by invitation)</td>
<td>School achievement and progress data, research milestone reports, classroom observation analyses</td>
<td>Termly (four per year)</td>
</tr>
<tr>
<td>Whole cluster hui</td>
<td>School leadership teams, representative teachers from each school, parent representatives, philanthropic funders, student representatives</td>
<td>Coherence, identifying innovation, review, and forward planning</td>
<td>Feedback from students, teachers, and researchers</td>
<td>Student, teacher, and researcher presentations</td>
<td>Yearly (August)</td>
</tr>
<tr>
<td>School PLCs</td>
<td>School staff</td>
<td>Implementation, pedagogical shift</td>
<td>Data discussions, knowledge building, professional learning and development</td>
<td>Targeted professional learning, school achievement data</td>
<td>Varied by school—typically twice termly (eight per year)</td>
</tr>
<tr>
<td>Team PLCs</td>
<td>Teaching teams</td>
<td>Planning and evaluation of practice</td>
<td>Planning, shared problem solving</td>
<td>Team-level achievement data, class and team sites</td>
<td>Typically fortnightly</td>
</tr>
</tbody>
</table>
Classroom Observations

Teachers were observed each year. To develop a context-specific observation tool to understand patterns of teaching, in the initial year a running record format was used by the lead authors to allow for open-ended, qualitative descriptions of classes, including specific descriptions of teacher activity. These were analyzed thematically, and from these naturalistic descriptions, a time sampling observation schedule was developed that included a precoded list of teacher-focused variables but allowed open-ended descriptions of the nature of student activity. As patterns of task provision also began to emerge, this tool was revised at the end of the second year to include precoding of the tasks students were engaged in when independent of the teacher, and the range of sites students accessed. The tool was designed to be descriptive, and was not specific to a curriculum area, but was designed to capture aspects of the shared pedagogy in the schools.

After the initial year, teachers were observed twice, in the third and ninth months of the school year. In each observation, 12 samples of 3-minute intervals were captured, wherein the observer alternated between observing the teaching, and noting the independent activities. In the 3-minute interval focused on the teaching, judgments were made about the nature of the main teaching activity (question and answer; lecturing or modeling; extended discussion or conferencing; roving; and management). Instances of feedback were recorded as evaluative, descriptive, and/or generative (feed-forward). Finally, all observed teaching foci were coded as item teaching (e.g., content terms), activating prior knowledge (e.g., reminding students of previously learned material), practice (e.g., applying a taught skill), strategy instruction (e.g., approaches that might support success), and critical thinking/literacy (e.g., text logic or positioning).

In the alternating 3-minute interval when the researcher observed any independent activities, researchers noted the nature of the digital sites used, the activities engaged in, and any student decision making. Collaboration was recorded as being through face-to-face (FTF) discussion, or a computer-mediated discussion (CMD) about individual (my work) or co-authored products (our work). Finally, activity during this interval was judged to have been managed offline (i.e., orally); online with some verbal teacher prompts; or totally digitally managed.

Data Analysis

e-asTTle Assessments

Repeated measures using students’ overall scores from the e-asTTle tool provided descriptions of achievement patterns over time. Student achievement data for reading and mathematics were matched to schools’ demographic roll data over 3 years (2012–2014), and over 4 years for writing (2012–2015). Because of revisions to the e-asTTle tool and subsequent renorming, writing achievement data collected at the initial testing point are not comparable, and have therefore been excluded from analyses.

National expectations provided in the e-asTTle assessment tool comprised different national rates expectations of levels and progress for different year levels. For each testing tool (writing, reading and mathematics) at each time point, overall scores were adjusted by normative averages with respect to year levels and time of testing. Achievement at the initial testing point was set as the cross-sectional baseline, representing change in achievement between year groups at the beginning of the study. Independent samples t-tests showed no significant differences in achievement between students who were tracked and those not tracked (because they left) at baseline for all testing tools.

The longitudinal analyses for students who remained in the intervention included summary of overall scores over time in comparison with predictive baselines and normative averages over three years (2012–2014). Students who were tracked through all time points were divided into five exclusive groups in longitudinal analyses (see Table 2).

Estimates of Effects

Estimates of growth across subgroups were undertaken through hierarchical linear modeling. The data set for the models comprised all students with data at two or more points any time during the
study. In the fourth year (2015), the schools decided to change the assessment tool used to monitor reading and mathematics; however, the assessment tool for writing remained unchanged. Thus, there were six time points of data for reading and mathematics: three full years. Because of the lack of data for the first testing point at the beginning of the study for writing, an additional year’s student achievement data was added to the multilevel model for writing, ensuring three full years of data in the model (seven time points, 2012–2015).

Three multilevel HLMs, for reading, writing, and mathematics, were used to estimate the growth trajectories. Overall growth patterns and variability between the levels of the schooling structure were estimated by the generic model and it then became the basis of comparisons for any added explanatory variables. The lmer() function in R (Bates, Maechler, Bolker, & Walker, 2015) was used for the modeling, and the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and comparison of the log likelihood were used to evaluate the goodness-of-fit between models.

The factors included in the models were student, gender, ethnicity, year level, school, and overall score at each time point. Using these data, a number of extra variables were calculated for each student, including a measure of “dosage” (i.e., number of times each student was in the respective databases); beginning year level; beginning school; gain score; difference from norm gain (i.e., difference between the student’s gain and the gain made at the national level for the same year levels and time periods); and the student’s initial score, relative to expected score for the same year level and time point. The difference from norm gain was the dependent variable for each model.

HLMs were fitted, with beginning school as the grouping factor. School explained very little of the variance within the data (only 2–4%). Models were fitted by sequentially adding explanatory variables, using analysis of variance (ANOVA) to determine significant differences between the two models, and the AIC and BIC to determine the best fitting model. Where an added variable was found not to add any explanatory power to the model, it was dropped, to ensure that the final model was the most parsimonious.

**Results**

**Longitudinal Cohort Tracking**

In Figures 1–3 students’ achievement scores over 3 years (2012 to 2014) are presented as visual comparisons with initial baseline levels and with test norms. In writing, tracked students made gains at rates in excess of national norms, with all cohorts on average approximating normative levels by the end of the third year. The overall pre–post average shift was 79 e-asTTle points more than the expected shift over 3 years. For all those students with complete data over the study (N = 553), the pre–post effect size shift in difference from norm was $d = 0.62$ ($t = -16.2$, $p < 0.001$). This effect can be considered a “moderate/high” effect by Cohen’s (1988) conventions.

In reading and in mathematics (see Figures 2 and 3), students exceeded the baseline levels set by previous cohorts in all cohorts but one, those students tracked from Year 7 through to Year 9, but no cohorts reached normative levels. In reading, for tracked students (N = 486), the average pre–post shift was 8.8 e-asTTle points in addition to normative shift, with an effect size $d = 0.13$ ($t = -3.4$,

---

1The New Zealand school year starts in late January and ends in December.
Figure 1. Average test scores for students tracked for 3 years in writing.

Figure 2. Average test scores for students tracked for 3 years in reading.

Figure 3. Average test scores for students tracked for 3 years in mathematics.
In mathematics, for tracked students \((N = 486)\) the average shift was 9.4 points in addition to normative shift, an effect size of \(d = 0.14\) \((t = -3.7, p < .001)\). In mathematics, effect sizes varied substantially by year level, with the largest effect for younger cohorts \((d = 0.53\) for those students tracked during Years 4–6), with effects reducing to negative effects in relation to norms for those students who moved into high school (see Figure 3).

**Estimates of Effects**

To account for the nested nature of the data and the variation in the students enrolled each year, HLMs were used to estimate the effects for students of varying levels of participation and in different schools within the cluster.

**Writing**

Estimates of the effects were calculated for all students present for two or more time points \((N = 2096)\) over the four years (2012–2015). For writing, the most parsimonious model contained all the explanatory variables except students’ beginning year level, which indicates similar effects across year groups. The full model suggests that, on average, students with any two or more time points of data made less than normative progress across the time periods. Boys’ gains were significantly smaller over the 3 years, averaging 29 points, or approximately three terms’ learning less than national norms for boys and girls combined \((t = 6.24, p \leq .001)\). However, the model also suggests acceleration as an effect of “dosage,” in that each additional time point was associated with an additional 7.5 e-asTTle points in addition to normative gain \((t = 5.48, p \leq .001)\). This would suggest that after their initial year, students achieved a cumulative gain of almost one term’s learning in addition to normative progress for each additional half year at the school. Students whose initial scores were relatively higher made smaller gains than those who began at lower initial levels of achievement, indicating a slight benefit for lower achieving students \((est = 0.56; t = 38.48, p \leq .001)\).

**Reading**

Estimates of the effects for reading were calculated for all students present for two or more time points \((N = 1694)\). The model that included all explanatory variables was selected as the most parsimonious model. The results from this analysis suggest that across the 3 years, students with data at any two or more time points on average made smaller than norm gains, but that each additional time point within an intervention school was associated with an additional gain of 5.22 e-asTTle reading points above normative gains \((t = 5.47, p \leq .001)\). This suggests that after a year in the schools, each additional half year equated to approximately half a term’s extra gain. Students who began at higher levels made smaller gains \((est = -0.13, t = -6.17, p \leq .001)\). On average, each increase in year level was associated with an increase in gain of 3 e-asTTle reading points greater than norm progress expectations \((t = 2.89, p \leq .01)\), with the largest gains being made by students in Years 5–6 (Year 5, \(n = 282, -x = 1\); Year 6, \(n = 269, -x = 11\)).

**Mathematics**

The most parsimonious model for mathematics excluded time in the cluster, schools, and gender. Thus, unlike for writing and reading, the scores between boys and girls and those who received different “dosages” were not significantly different. Overall, for all students who were present at two or more time points, the average gain was 47 points (approximately an extra year’s learning) higher than norm gains across the 3 years. Students who began closer to norm levels made smaller gains than those that began at lower initial levels of achievement \((est = -0.20, t = -10.91, p \leq .001)\). As with the other analyses, students within different schools made differential gains. Unlike the other subjects, each increase in beginning year level was associated with a decrease in gain, relative to the norm. Students who entered our databases at Year 6 or below made greater gains over the time periods relative to the norm, while students in Years 7–10 made smaller gains than the norm \((est = -9.04, t = -9.59, p \leq .001)\).
Pedagogical Effects
The refinements to classroom instruction were investigated using classroom observations. These observations showed continual high levels of digital implementation, with more than 80% of students using their devices in a lesson, in more than 86% of observed classes, and students accessing learning digitally (through the class site) in 85% of observed intervals. As part of the iterative process, results of observations were fed back to teachers, discussed in relation to the shared hypotheses, and further refinements were made. As discussed previously, patterns of teacher interaction were categorized across 2 years; patterns of student activity were categorized in the final year.

Teaching Patterns
Observations of main teaching activities indicated shifts in teachers’ interaction styles consistent with Hypothesis 4: refinements that focused on supporting deeper thinking (see Table 3). Through discussion with teachers, we expected that higher rates of conferencing and extended discussion, and lower rates of question and answer (Q&A) sequences and roving to assist students while working, would be most likely to foster deep thinking. In the initial observation, teachers spent approximately a quarter (28%) of observed intervals in conferencing or extended discussion. Similar numbers of intervals were spent in Q&A sessions (23%), and roving (26%). By the final observation point, half of all observed instructional intervals (50.6%) were spent in extended conversations or conferencing, with less time in Q&A sequences (13%) or roving (13%). An unanticipated finding was that more time was given to modeling or lecturing over the time. This pattern may also be consistent with a shift toward cognitive complexity, but was not an explicit expectation in data discussions.

As well as changes in the interaction style, the focus of the interactions was also observed to shift. For this variable, observers noted any instance of an identified teaching focus during observed intervals. While all teaching foci are an important contribution, through data discussions, the relative complexity of these foci was problematized. In the initial observation, teachers included practice of taught skills in 42% of observed intervals, and strategy teaching in 25% of intervals. These rates stayed relatively stable. However, over time, rates of activating prior knowledge rose (from 4.4% to 31.6%), as did rates of content item teaching (from 3.7% to 26.4%). Rates of teaching critical thinking were very low in the first observation (observed in 2.2% of intervals). Shifts in instruction consistent with Hypothesis 4 might be expected to show increases in rates of critical thinking. While rates rose markedly, they maintained comparatively low overall (from 2.2% to 7.5% of intervals). Overall patterns showed teachers including a wider range of instructional foci, but continued lower levels of critical thinking.

Levels and types of feedback observed and patterns of collaboration also shifted over time. In the initial time point, feedback was observed in almost half of all observed intervals (49.6%); by the final observation, rates had increased to 86% of intervals. Moreover, the types of feedback had shifted, to include greater incidence of generative feedback (sometimes called “feed-forward”), from 12.5% of observed intervals to 26% of observed intervals. Peer collaboration rates also rose, from being observed in 55% of intervals, to 72% of intervals.

Student Activity
In the final year, student activity was coded, and shifts noted in line with Hypothesis 3: increasing complexity. In general, more constrained tasks were assumed to indicate lower expectations of

---

Table 3. Main Teaching Approach Taken in Observed Intervals

<table>
<thead>
<tr>
<th>Time 1 (135 intervals)</th>
<th>Time 2 (134 intervals)</th>
<th>Time 3 (174 intervals)</th>
<th>Time 4 (174 intervals)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference/ED</td>
<td>28.1</td>
<td>16.4</td>
<td>39.8</td>
</tr>
<tr>
<td>Lecture/model</td>
<td>2.2</td>
<td>6.0</td>
<td>6.4</td>
</tr>
<tr>
<td>Q&amp;A</td>
<td>23.0</td>
<td>33.6</td>
<td>17</td>
</tr>
<tr>
<td>Instructions</td>
<td>13.3</td>
<td>19.4</td>
<td>9.4</td>
</tr>
<tr>
<td>Roving</td>
<td>25.9</td>
<td>23.1</td>
<td>21.1</td>
</tr>
<tr>
<td>Management</td>
<td>0.7</td>
<td>0.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Others</td>
<td>6.8</td>
<td>1.5</td>
<td>5.1</td>
</tr>
</tbody>
</table>

---

Jesson et al.
complexity, whereas more open tasks, such as creation of digital learning objects (DLO), extended writing, and extended reading, were assumed to include greater opportunity for complexity. Figure 4 shows the categorized student activity, indicating slight decreases in the numbers of constrained practice activities observed, and increases in more open-ended activities, for example, creation of DLO, writing, and reading. Patterns of student activity also showed continued relatively low levels of using multiple texts in lessons.

Discussion

A research–practice partnership was used to refine an educational intervention aimed at improving students’ achievement outcomes through a digital learning initiative. The pedagogy was designed to promote student engagement through students’ creation of digital artifacts about their learning, which were posted onto their blogs, thereby capitalizing on features of the digital environment, including productive activities for students’ knowledge creation (McLoughlin & Lee, 2008) and increased engagement (Bebell & O’Dwyer, 2010). Research findings as feedback throughout the intervention enabled the schools to investigate the extent to which their efforts constituted an improvement. Four hypotheses drove the partnership activities, arising from a sociocultural framing of development, the profile identified at the initial phase of the study, and negotiated through PLCs with teachers and leaders. Overall results indicate some ongoing shifts in pedagogy in line with shared hypotheses, alongside improvements of varying degrees for learners, as follows.

Hypothesis 1: A Focus on Accelerating Student Achievement

A key activity of the RPP was formative use of student achievement data to monitor ongoing improvement. Effects on progress greater than normative expectations were found in writing, reading and mathematics for students who stayed in the intervention for 3 years, with the largest effect size in writing. The increased effect on writing for all year levels in this study is consistent with the Zheng et al. (2016) meta-analysis of effects internationally, which identified the greatest and most consistent effect in the area of writing; however, the effect size shift for students over the 3 years was greater than that reported by Zheng et al., whereas effect sizes for reading and mathematics were comparable.

While the longitudinal analyses identify effects to varying degrees for students who were at the schools over the full 3 years, our analyses also estimated the effects for students involved to varying degrees. These analyses reflect more closely the state of affairs operating in schools in low-income communities with changing populations, continually shifting the makeup of classes and the presenting profile of strengths and needs for teachers to accommodate. In reading and writing, there was an effect of “dosage,” indicating that only with increased time in the intervention (three time points—a
full calendar year) was there any associated accelerated literacy learning for students. A limitation of this finding is that these gains may be related to stability in general, as distinct from participation in digital learning in particular. Effects in mathematics were also apparent, but were not subject to “dosage” in the same way, and did not continue into the higher year levels.

Hypothesis 2: Growing Effective Practices
In line with the features of RPPs, variability is seen as an important source of knowledge (Snow, 2015). One tentative explanation for the noted variability between learning areas relates to the activity within the RPP. While student achievement data were tracked in all areas, and digital pedagogy was a ubiquitous feature of the teaching in all subjects, the focus of the research into effective pedagogy in Phase Two was writing. Here, the form of variability we capitalized on was the variability between teachers. Specifically, the case studies of teachers enabled the partners to identify practice associated with demonstrably effective teachers in writing. Thus, shifts in pedagogy in line with these case-study teachers would be expected to impact most in the area of writing.

Arguably, therefore, small shifts in the outcomes for reading and mathematics occurred alongside the shift to ubiquitous digital pedagogy, without as intense a research and development focus on pedagogy in those areas. In addition, while achievement patterns were analyzed in all tested areas, and observations were conducted generally, the improvement activities of the cluster were focused on the underperforming area of writing. Thus, attention to refining practice in line with innovation within the cluster was geared toward pedagogy in writing, although it was not intended to preclude shifts in achievement in reading and mathematics. In line with the enduring nature of partnerships (Snow, 2015), and based on the findings of this initial study, ongoing problems of practice have since been identified for closer investigation, including a potential plateau in reading, limited shift in mathematics over time, and a brake on achievement over summer.

Hypothesis 3: Increasing Cognitive Complexity of Activities
Like studies of digital initiatives internationally, our initial profile suggested that students were largely engaged, in that they were on task, and reported an affinity with school and the goals of the digital initiative (Jesson et al., 2015). The focus for refinements therefore was to increase students’ cognitive engagement, through activities and tasks with greater complexity. Classroom observations indicated that students were increasingly engaged in more open-ended tasks, including creating DLOs, and extended reading or writing, whereas they were less often engaged in practicing skills through worksheets or learning games.

One assumption underpinning the shared pedagogy of creation activities was that reading of multiple texts, and synthesis through creating, would increase the cognitive complexity of independent activity. Thus, “creative” activity would incorporate originality of thinking and synthesis of ideas, as well as multimodal presentation, and blogposting would provide students opportunities to create an online presence as a “producer” of knowledge (McLoughlin & Lee, 2008). Within this conception also lies a potential explanation for effects on writing. The posting of one’s creations, conceived as authorship, may have important impacts for writing. It may be, for example, that productive activities, whether multimodal or solely written, develop skills assessed within a writing test: an awareness of audience, and of organization of text, structured and logical development of ideas, use of sentence variety, and precise choices of vocabulary. However, it was also the case that, as in other studies of digital implementation (Zheng et al., 2016), students spent more time writing.

Hypothesis 4: Shift to More Interactions Extending Deeper Thinking
As with other studies of digital initiatives, our study identified opportunities for increased leverage from enhanced teaching interactions (Lowther et al., 2012). Shared hypotheses for improvement were premised on initial profile of efficient classes, and therefore available teacher time to focus on moving thinking forward through face-to-face discussion (Jesson et al., 2015). Interaction
patterns shifted in line with increased expectations for students to think deeply. Patterns of short Q&A sequences decreased, and extended discussions increased. Teachers also increased the amount of feedback to students, as well as the opportunities for students to collaborate. Alongside these shifts were student achievement gains, most apparent in writing. It may be that this impact on writing reflects opportunities within the digital environment for more effective feedback or collaboration to be more readily leveraged in writing than either reading or mathematics, through, for example, co-authorship and feedback during the creation of artifacts for blogposts.

Although teachers were successful in redesigning some aspects of pedagogy, other aspects were identifiable as more difficult to change, with the classroom observations indicating less shift in practice specifically in the number of intervals where critical thinking was the identified focus. Arguably, the ability to engage critically with material is crucial to successful learning in a digital environment, and numbers of intervals in which students were asked to think critically rose threefold, but remained relatively low. This may reflect a relative difficulty for teachers of leading conversations that support these higher order thinking skills, identified as not well taught generally (Grimes & Warschauer, 2008). Similarly, synthesis and evaluation require the reading and analysis of multiple texts; however, this activity was not common in our observations. More common was students writing or creating based on the reading of a single text, arguably less complex, but in line with known forms of literacy instruction. Both critical thinking and multiple text use are highly complex forms of instruction, and may indicate a need for targeted development to support teachers to plan for critical thinking interactions and activities that required synthesis and evaluation across texts.

**Limitations**

We have focused here on the outcomes over time of the ongoing research practice partnership in terms of student achievement shift and teacher practice. Accordingly, the study is limited in its ability to explain the causes of the reported effects. Undeniably, a myriad of variables, not measured or reported here, will impact students’ learning and the variation of implementation of the initiative, including but not limited to parental engagement, professional development, teacher attrition, teacher knowledge, and leadership practice. Moreover, to gain a detailed understanding of the nature of the pedagogy, qualitative data are needed to understand the nature of the changes in a more detailed and nuanced way. Avenues for further research, therefore, will include qualitative investigation into how the digital pedagogy is understood and implemented by the teachers and leaders, alongside insights from the students’ and parents’ perspectives.

**Conclusion**

This study has documented the impact of a research–practice partnership in which research findings were used formatively by schools to strengthen a digital innovation. Like other digital initiatives, effects were largest in the area of focus for the schools, writing; however, there were discernable effects in terms of progress in reading and mathematics. The results indicate some support for our hypotheses: that accelerated gains could be made if teachers and schools focused on identifiable opportunities in the digital environment by using the shared pedagogy incorporating creation activities and blog posting, to design higher order activities, and to focus interactions on deepening thinking. While there were shifts toward patterns of classroom activity and interaction in accordance with these hypotheses, future research needs to understand how to best use the opportunities of the digital environment for large-scale traction in reading and mathematics.

Received 14 June 2017
Initial decision 24 January 2018
Revised manuscript accepted 30 January 2018

**Funding.** The research reported here was co-funded by the Education Trust of the participating schools and the University of Auckland.
Author Notes

Rebecca Jesson is a Senior Lecturer at the Faculty of Education and Social Work at the University of Auckland. She is Associate Director of the Woolf Fisher Research Centre, where she works with teachers and school leaders to redesign their instructional provision based on research evidence of what is working for their particular context. Rebecca’s research interests are in the teaching and learning of literacy, and use of dialogic pedagogies to promote learning. Please address correspondence regarding this article to Rebecca Jesson, Woolf Fisher Research Centre, Faculty of Education, University of Auckland, Private Bag 92601, Symonds Street, Auckland 1150, New Zealand. Email: r.jesson@auckland.ac.nz

Stuart McNaughton is Professor of Education in the Faculty of Education and Social Work. His research programme includes work on the development of a psychological model of socialisation (incorporating concepts of teaching, learning and development) applicable to informal and formal educational settings which provides a means of analysing development within and across settings. Associated with this is the demonstration through research applications of ways of incorporating cultural processes in research tools and in explanations of teaching, learning and development. These applications contribute to solutions to a long standing difficulty in developmental and educational psychology, explaining the role of culture in teaching and learning.

Aaron Wilson (PhD) is a senior lecturer in the School of Curriculum and Pedagogy and an Associate Director of the Woolf Fisher Research Centre. His research interests are in interventions to address disparities in education, disciplinary literacy teaching in secondary schools, English, blended learning, and teacher professional development. Aaron has led the literacy strand of the Starpath project since 2011 and was the Principal Investigator and National Coordinator for the Secondary Literacy Project 2009-11, an intervention in 60 secondary schools that aimed to raise the achievement of underachieving Year 9 and 10 students.

Tong Zhu is a research fellow at the Woolf Fisher Research Centre at the Faculty of Education and Social Work at the University of Auckland. His current research interests include the analysis of hierarchical structured data (with particular application to New Zealand primary and secondary student achievement data), categorical data analysis, applied statistics and data visualisation.

Victoria Cockle is a research analyst at the Faculty of Education and Social Work at the University of Auckland. Her research interests include mixed methods analyses which contribute empirical understanding to projects which work to better meet the needs of underserved students in New Zealand and in Pacific Nations.

References


Jesson et al.


