PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

EVALUACIÓN DEL USO DE MULTIDISPOSITIVOS Y SHARED DISPLAY COMO HERRAMIENTAS PARA LA ENSEÑANZA COLABORATIVA DEL LENGUAJE EN ALUMNOS DE TERCERO BÁSICO

TOMÁS IGNACIO MARTÍNEZ GALARZA

Tesis para optar al grado de
Magíster en Ciencias de la Ingeniería

Profesor Supervisor:
MIGUEL NUSSBAUM VOEHL

Santiago de Chile, (Diciembre, 2015)

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

EVALUACIÓN DEL USO DE MULTIDISPOSITIVOS Y SHARED DISPLAY COMO HERRAMIENTAS PARA LA ENSEÑANZA COLABORATIVA DEL LENGUAJE EN ALUMNOS DE TERCERO BÁSICO

TOMÁS IGNACIO MARTÍNEZ GALARZA

Tesis presentada a la Comisión integrada por los profesores:

MIGUEL NUSSBAUM VOEHL

YADRAN ETEROVIC SOLANO

CARLOS GONZÁLES UGALDE

MIGUEL ATTILIO TORRES TORRITI

Para completar las exigencias del grado de

Magíster en Ciencias de la Ingeniería

Santiago de Chile, (Diciembre, 2015)

A mi familia, sin su apoyo constante esta tesis habría sido olvidada.

A mis amigos, por su fe incondicional.

AGRADECIMIENTOS

A mis padres, quienes han sido siempre un ejemplo a seguir y cuyo esfuerzo me ha permitido tanto.

A todos mis hermanos y hermanas, que han estado siempre presentes en mi vida y escucharon hablar de este proyecto durante mucho tiempo.

A mis amigos que siempre me han apoyado, y sin los cuales las pruebas de las distintas aplicaciones no habrían sido lo mismo.

A Andrea Vásquez, Enzo Sciarresi, Gabriel Vidal, Esteban Serrano y todos los otros miembros del proyecto, jamás podríamos haberlo logrado sin ustedes.

A toda la comunidad del Departamento de Ciencias de la Computación, quienes nos acompañaron en todas las etapas y pruebas del sistema, apretando teclas y escuchando las mismas frases una y otra vez.

Al profesor Miguel Nussbaum, a quien ni la convalecencia ni la distancia le impidió preocuparse por el avance de mi tesis y sin su apoyo y comprensión no estaría donde estoy hoy.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

RESUMEN

Multidispositivos y shared display son tecnologías que permiten, respectivamente, utilizar un solo computador con muchos dispositivos de input (teclados) distintos en simultáneo y una misma pantalla proyectada para múltiples alumnos.

Al combinar estas tecnologías se logra generar sistemas para la enseñanza que simulan que cada alumno tiene su propio computador a una fracción del precio. Además, al estar todo conectado en un solo computador se abren nuevas oportunidades, específicamente la posibilidad de realizar trabajo en grupos en vez de individual, de donde nace la pregunta: ¿Es posible desarrollar colaboración en grupos pequeños con shared display para enseñanza del lenguaje?

Con este fin se desarrolló software para evaluar las capacidades ortográficas de niños de tercero básico en base a un proyecto similar para matemáticas existente (MultiMouse), que al fin de cada sesión genera un reporte de las áreas más débiles de los alumnos y así se realiza un repaso dirigido. Este proceso fue realizado a lo largo de un semestre en un colegio con alumnos de tercero básico.

Los resultado obtenidos no muestran una mejora significativa respecto a los resultados obtenidos con clases tradicionales, y esto se considera que se debe a varios factores, entre ellos la inestabilidad física del sistema y las condiciones poco favorables del colegio con el que se trabajó. Sin embargo, la arquitectura de software desarrollada para el sistema colaborativo fue reutilizada exitosamente para un paradigma individual, con el que se condujo otro experimento en donde se demostró la diferencia que distintos tipos de tecnología tienen en la eficacia de la introducción de tecnología en la sala de clases.

Esta investigación cuenta con el apoyo del Centro de Estudios de Políticas y Prácticas en Educación, Beca CIE01-CONICYT.

Palabras claves: Single Display Groupware, Computador Interpersonal, Shared Display, Multidispositivos.

ABSTRACT

Multi-devices and shared display are technologies which allow us, respectively, the use of a single computer with several input devices (keyboards) simultaneously and a single projected screen for multiple students.

Combining these technologies makes systems where each student is given the experience of a single computer at a fraction of the cost possible. Also, having the system running on a single computer opens up new opportunities, specifically a system where students work in groups, instead of individually. From these ideas arises the question: Is it possible to use shared display to generate collaboration in small groups as a tool for language teaching?

To this end software was developed to evaluate orthography in third grade students based on a similar existing mathematics project (MultiMouse), and at the end of each session a report was generated highlighting the weakest subjects and an oriented review was conducted with the students. This process was testes throughout one semester in a school with third grade students.

The results obtained do not show a statistically significant improvement over scores results obtained via standard teaching methods, this was considered to be due to several factors, among them the real world instability of the system built along with the less than favorable conditions in the school where the test was conducted. However, the software architecture developed was successfully reutilized in an individual paradigm, with which an experiment was conducted that proved the different levels of efficiency that different types of technology have when introduced to a classroom.

The research was supported by the Center for Research on Educational Policy and Practice, Grant CIE01-CONICYT.

Keywords: Single Display Groupware, Interpersonal Computer, Shared Display, Multi-devices.

I. CAPÍTULO 1: USO DE MULTIDISPOSITIVOS Y SHARED DISPLAY COMO HERRAMIENTAS PARA LA ENSEÑANZA COLABORATIVA DEL LENGUAJE EN ALUMNOS DE TERCERO BÁSICO

1. INTRODUCCIÓN Y MOTIVACIÓN

La situación actual de la educación chilena es que en las escuelas públicas un curso promedio tiene alrededor de 45 alumnos, los cuales son responsabilidad de un único docente, que usualmente no tiene recursos suficientes para mantener a los alumnos involucrados en la clase y la experiencia de los alumnos no está optimizada para cada uno.

La UNESCO destaca que para la enseñanza de la escritura es necesario generar indicadores de desempeño desagregados para cada ámbito de la evaluación de la escritura (UNESCO 2010). Uno de los elementos más complejos de aprender de la escritura del lenguaje español es la ortografía, ya que el lenguaje posee muchas particularidades. Debido a esto se debe plantear una didáctica específica para su enseñanza (Sotomayor, Molina, Bedwell, & Hernández, 2013).

Una forma de enseñanza que busca resolver estos problemas es el *Task based approach*, cuyo objetivo es crear “condiciones naturales” que permitan desarrollar competencias lingüísticas a nivel gramatical y de adquisición de significado, a través de involucrar al estudiante en procesos mentales que facilitan la expresión de estas competencias lingüísticas en tareas (Prabhu, 1987). La forma de aplicar la tecnología para lograr estas condiciones se conoce como *Computer Assisted Language Learning* (CALL), un enfoque que busca usar la tecnología para presentar y reforzar material educativo considerando elementos de interacción (Bax, 2003).

Cabe destacar también que se ha observado que los docentes que son capaces de utilizar una variedad de estilos de interacción, en lugar de un acercamiento rígido a la enseñanza tienden a ser más exitosos (Darling-Hammond 2000).

El paradigma de aprendizaje colaborativo plantea que es posible que los alumnos aprendan en conjunto, como un grupo, a través de interacciones sociales entre ellos (Vygotsky, 1979; Rogoff, 1994). Cuando se utilizan computadores y tecnología para generar el ambiente con el que los alumnos van a trabajar colaborativamente se habla de *Computer Supported Collaborative Learning* (CSCL) (Roschelle & Teasley 1995, Salomon 1990). El CSCL se ha utilizado en repetidas ocasiones, con positivos resultados tanto a nivel de interacción entre alumnos como a nivel de aprendizaje del material estudiado (Scott, Shoemaker and Inkpen, 2000; Jipping, Dieter, Krikke & Sandro, 2001; Stanton, Neale & Bayon, 2002; Danesh, Inkpen, Lau, Shu & Booth, 2001).

La forma elegida en esta tesis para aplicar tecnología y aprendizaje colaborativo a la enseñanza de la ortografía del lenguaje español combina el uso de *Shared Display* y Computador interpersonal, tecnologías que se han usado con anterioridad para el exitoso aprendizaje de aritmética en alumnos (Alcoholado, Nussbaum, Tagle, Gomez, & Denardin, 2012). Con el desarrollo de la aplicación "*One Mouse Per Child for Basic Math*" se lograron resultados estadísticamente significativos en el aprendizaje de la adición, calidad del trabajo realizado y porcentaje de logro en la actividad respecto a un pre-test aplicado.

Es en base a estos principios y tecnologías que se plantea la siguiente pregunta de investigación: ¿Es posible desarrollar colaboración en grupos pequeños con shared display para enseñanza del lenguaje? Para responder esta pregunta se ha desarrollado el sistema descrito más adelante, el cual funciona como herramienta de enseñanza y una fuente de información para los docentes, los cuales obtienen información respecto al nivel de los alumnos y con esto pueden concentrarse en las áreas que requieren mayor atención.

2. SISTEMA

El sistema desarrollado está construido bajo dos principios generales, Computador Interpersonal y Shared Display. Computador Interpersonal consiste en permitir que diversas personas ocupen un mismo computador en simultáneo, a través de introducir múltiples sistemas de input de usuario que operan en paralelo. Shared Display es una idea semejante, pero aplicada a la pantalla de un computador, en donde la pantalla es compartida por múltiples usuarios. Usualmente para el uso de Shared Display se debe trabajar con un proyector, ya que esto permite que todos los usuarios tengan buen acceso a la interfaz de usuario.

Se determinó que para la enseñanza de la ortografía el uso de mouse era insuficiente, por lo que para este sistema se utilizaron teclados y audífonos, de forma que cada alumno tenía asignado un teclado y un audífono, junto con un sector limitado de la pantalla. El uso de teclado permite presentarle al alumno múltiples tipos de problemas, desde selección múltiple a dictado de frases complejas, oportunidades que fueron aprovechadas con los diversos tipos de preguntas que se generaron para el sistema. Uno de los puntos más importantes que se buscó mantener de la investigación anterior (Mouse-Based Interpersonal Computer for Individual Learning; Alcoholado et al, 2012) fue el feedback instantáneo para el alumno, ya que este se determinó que es uno de los factores que más influyó en el éxito de dicha investigación.

Físicamente, el sistema desarrollado consiste de un computador local sin acceso a internet y un número igual de teclados y audífonos conectados a dicho computador, los cuales eran repartidos por la sala de clases mediante el uso de cables extensores de USB y hubs USB (Figura 2 – 1). Este sistema físico es el mismo utilizado para el sistema individual como para el sistema colaborativo.

Figura 2 – 1: Sistema físico armado en sala de clases.

La herramienta desarrollada para el trabajo colaborativo fue desarrollada en paralelo junto con una herramienta para el aprendizaje individual, la cual fue probada con alumnos en simultáneo y cuyo desarrollo fue dirigido por Enzo Sciarresi bajo la guía de la alumna de doctorado Andrea Vásquez y con colaboración del tesista (Sciarresi, Nussbaum 2014), pero ambos sistemas comparten una arquitectura. Este sistema individual estaba a su vez basado principalmente en el trabajo desarrollado para la enseñanza de aritmética con los principios de Shared Display y Computador interpersonal, Single Display Groupware (Alcoholado et al, 2012), el cual utilizaba múltiples mouse y una pantalla proyectada compartida por los alumnos.

Las preguntas a realizarle a los alumnos fueron construidas por un equipo de pedagogos de la Universidad Católica y representan los contenidos que los alumnos ven en tercero y segundo básico. Las preguntas se encuentran divididas en niveles, que representan tanto distintos contenidos como niveles de dificultad, estas preguntas son las mismas utilizadas en la experimentación del capítulo uno como del capítulo dos.

El sistema colaborativo es el foco del capítulo uno, mientras que el capítulo dos se concentra en el sistema individual y una comparación de este con un sistema basado en tecnologías distintas.

2.1 Single Display Groupware Individual:

El sistema individual es descrito en profundidad en el capítulo dos, sin embargo, para formar una imagen más completa, una descripción general sigue a continuación. La pantalla proyectada era dividida en una cantidad de secciones igual al número de teclados detectados en el sistema y a cada uno de estos sectores se les asignaba un color de fondo para ayudar a los alumnos a ubicarse y mantener el foco en su espacio asignado.

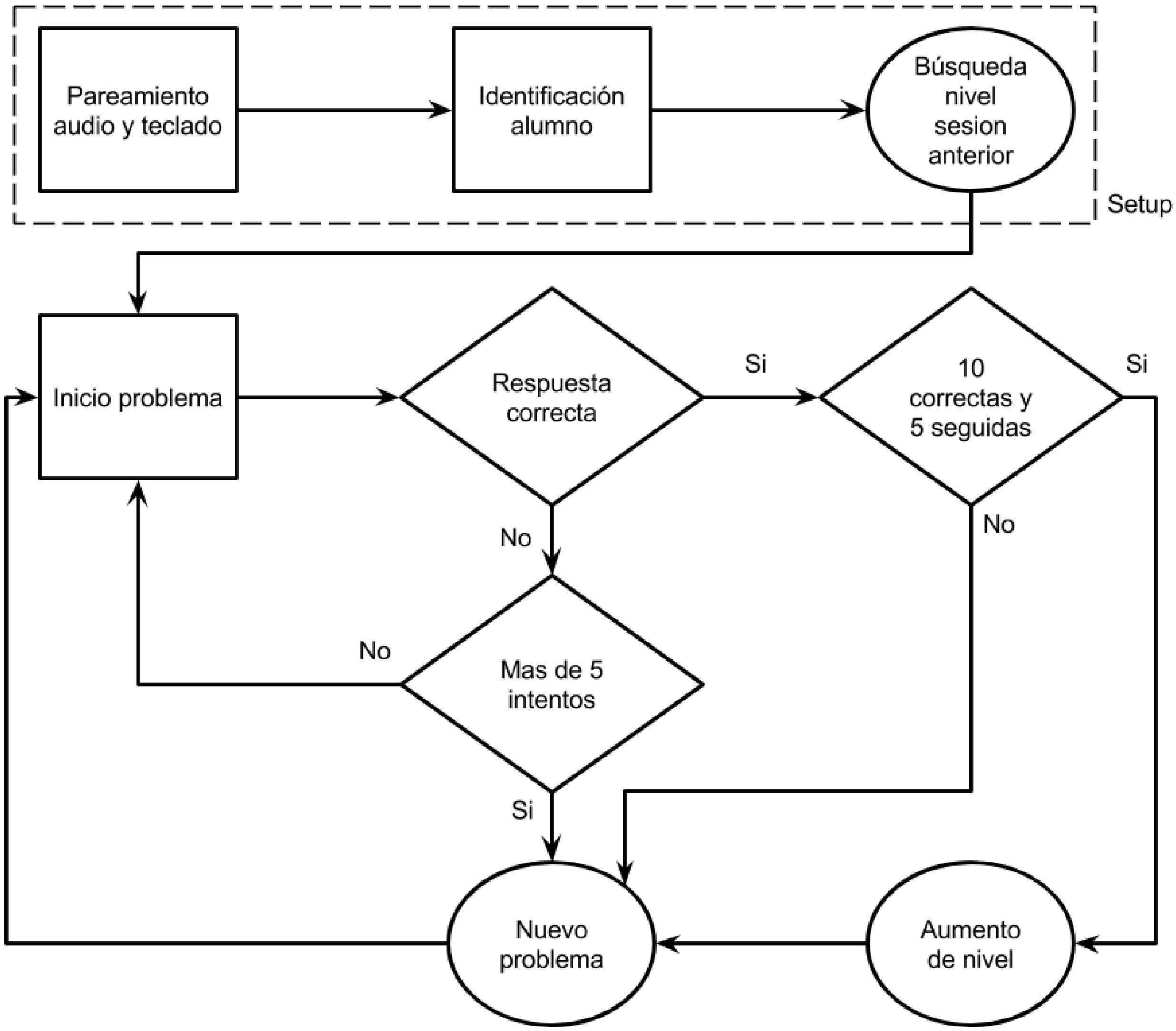

Figura 2.1 – 1: Flujo Single Display Groupware Individual

La interacción del alumno con el sistema comienza con el pareamiento del teclado y el audífono, ya que no es factible identificar a través de software que teclado está ubicado físicamente junto a qué audífono, se construyó un sistema mediante el cual los audífonos a los cuales no se les ha asignado un teclado repiten un número único a ellos. Si este número es ingresado en un teclado que se encuentra en la etapa de pareamiento (no tiene un audífono asignado) entonces se crea una asociación entre este teclado y ese audífono (Figura 2.1 – 1, primer cuadrado).

Luego de completar el pareamiento al alumno se le pide que ingrese su número de lista, con esto se puede identificar al alumno para el feedback y para que comience en el nivel en donde quedó en la sesión anterior. El número de alumno también es utilizado para el registro de avance y acciones de los alumnos (Figura 2.1 – 1, segundo cuadrado).

La sección de la pantalla asignada al alumno se encuentra divida en una sección de pregunta y una sección de respuesta, el tamaño de ambas siendo variable para adaptarse a preguntas más largas. La sección de respuesta consiste de un espacio para escribir (textbox) para las preguntas en donde al alumno se le dicta o se busca que escriba la palabra que falta o un sistema de selección múltiple, para las preguntas en donde se busca que el alumno elija la alternativa correcta (Figura 2.1 – 2).

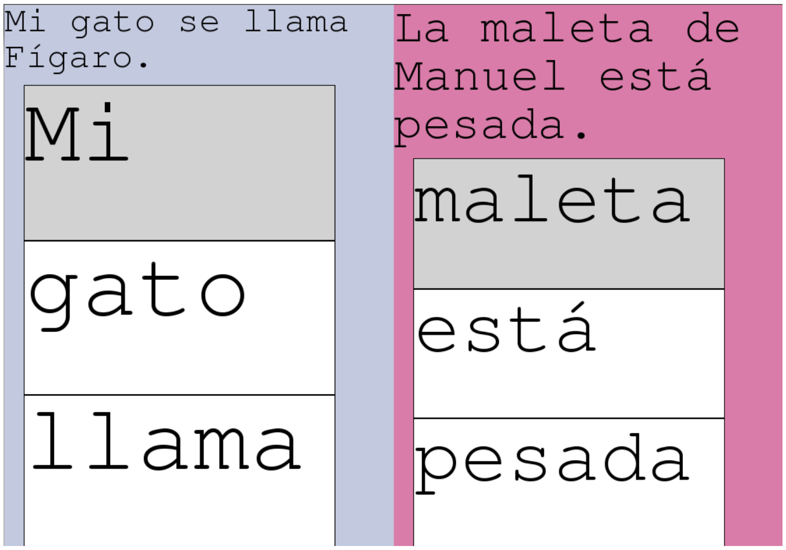

Figura 2.1 – 2: Sistema individual mostrando dos alumnos en pregunta de selección múltiple.

Para avanzar en el sistema, el alumno debe seguir las instrucciones que le son dadas a través de los audífonos, estas instrucciones corresponden a la pregunta que el alumno está trabajando en ese minuto y como responderla. En todo momento mientras no se esté ya reproduciendo audio, el alumno puede presionar la tecla Menú (Figura 2.1 – 3) en el teclado para repetir la última instrucción que le fue dada.

Figura 2.1 – 3: Tecla menú de un teclado estándar.

El alumno avanza de nivel de pregunta cuando ha respondido correctamente al menos diez preguntas del nivel actual y ha respondido al menos cinco preguntas correctas de manera consecutiva. Si un alumno se equivoca en la misma pregunta más de cinco veces, se le da la respuesta y se le hace otra pregunta del mismo nivel (Figura 2.1 – 1, sección inferior del diagrama). El funcionamiento y principios del sistema individual son descritos en mayor profundidad en el capítulo dos.

2.2 Single Display Groupware Colaborativo

Como se discutió previamente, la colaboración entre los alumnos es una de las mejores herramientas a la disposición de un docente para la enseñanza de materias complejas, como lo es el lenguaje español. Es por esto que se buscó aprovechar la oportunidad que presenta el computador interpersonal para el trabajo en equipo. Se buscó desarrollar un sistema que le permitiera a los alumnos recibir feedback inmediato pero al mismo tiempo aprender unos de otros, para esto se construyó el sistema descrito a continuación:

La herramienta desarrollada para el trabajo colaborativo de los alumnos está en gran parte basada en la herramienta individual y por lo tanto comparte muchas características con esta, sin embargo, la naturaleza colaborativa agrega un nivel de complejidad mayor tanto al funcionamiento del sistema como a la preparación de este. El sistema está diseñado para grupos de tres alumnos, ya que este número se consideró óptimo para generar discusión entre los alumnos pero a la vez lograr un consenso rápido (dado que los cursos no siempre tienen un número de alumnos múltiplo de tres se diseñó el sistema de manera que también permitiera grupos de a dos o hasta un alumno).

El flujo a seguir en una sesión de trabajo colaborativo consta de dos partes principales, las cuales son descritas a continuación:

2.2.1 Setup

A diferencia del sistema individual, en donde los espacios asignados a los alumnos desde un principio son los espacios con los que van a trabajar durante toda la sesión, en el sistema colaborativo es necesario formar los grupos, por lo que la pantalla desde un principio está orientada a el trabajo en grupo. Esto se logra dividiendo la pantalla en un número igual a la cantidad de grupos que hay y luego cada uno de estos espacios se divide en tres, asignándole así a cada alumno su espacio inicial para la configuración del sistema.

Este sistema es el foco de esta tesis, en el segundo capítulo se entra en mayor detalle respecto al sistema individual, sin embargo la arquitectura común de ambos sistemas es el principal aporte de esta tesis.

La primera etapa de operación del sistema es el setup (Figura 2.2.1 – 1), en donde se identifica la configuración física del sistema, se identifican a los alumnos y se forman los grupos. La pantalla inicial del sistema muestra un sistema para el pareamiento de teclado y audífono, en donde hay un número de espacios igual a la cantidad de alumnos que van a trabajar para que ingresen los números que escuchan en los audífonos, creando así la conexión entre teclado y audífono (Figura 2.2.1 – 1, primer cuadrado).

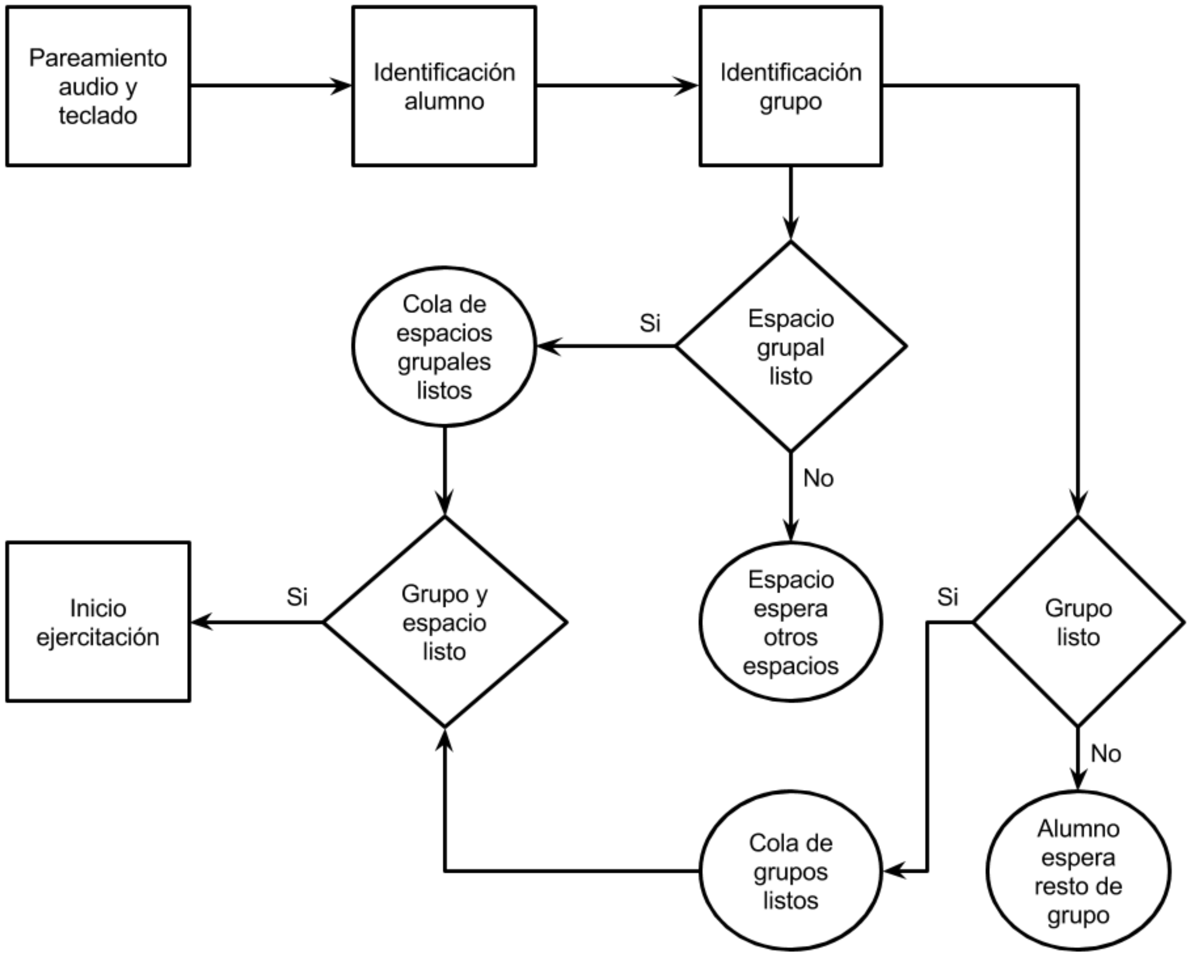

Figura 2.2.1 – 1: Flujo de setup colaborativo.

Luego, se procede a identificar al alumno y a su grupo (Figura 2.2.1 – 1, segundo y tercer cuadrado). Esto se hace pidiéndole al alumno que ingrese su número de lista y luego su número de grupo. Una vez que el alumno ha ingresado toda esta información se considera que el alumno está listo para comenzar y queda esperando a que el resto de su grupo esté listo.

Cada vez que un alumno termina la configuración inicial se revisa si es que todo su grupo está listo, en cuyo caso se agrega el grupo a una cola de grupos que están esperando que se les asigne un espacio grupal. A su vez, se marca el espacio individual del alumno como disponible y si los tres espacios individuales de ese espacio grupal están listos se agrega el espacio grupal a una cola de espacios grupales disponibles. De haber al menos un espacio grupal disponible y un grupo listo para comenzar se le asigna ese espacio grupal a ese grupo, sin importar si es que los alumnos antes se encontraban en este espacio o no. Esta configuración se debe principalmente a que no es posible identificar la ubicación física de los teclados y audífonos, por lo que la posición inicial de cada alumno en la pantalla es al azar, y dado que se quiere que los grupos estén formados por alumnos que se encuentran sentados juntos una vez que se ha logrado identificar los alumnos que forman un grupo se deben unir en un espacio disponible.

Figura 2.2.1 – 2: Display con grupos listos para trabajar.

Una vez que al grupo de alumnos se les ha asignado un espacio grupal se les informa por los audífonos que están listos para comenzar y que busquen su nombre y grupo en la pantalla. El display inicial para marcar que el grupo está listo para comenzar es simplemente el espacio con los nombres de los alumnos del grupo (Figura 2.2.1 – 2), de forma que ellos se puedan encontrar en la nueva posición y cuando los tres marquen que están listos para comenzar (presionando la tecla enter) se procede al primer nivel de ejercitación (Figura 2.2.1 – 1, último cuadrado).

2.2.2 Ejercitación

El nivel en el que cada grupo parte en cada sesión es el nivel del alumno del grupo que haya terminado su última sesión en el nivel más bajo, con el objetivo de que ningún alumno se pierda revisar alguna parte de los temas a aprender.

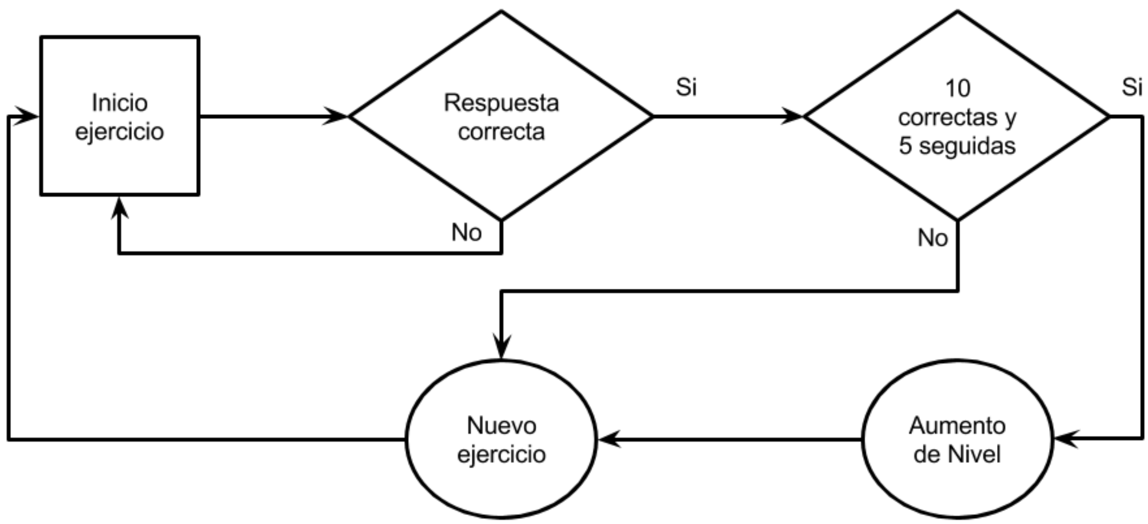

Figura 2.2.2 – 1: Flujo ejercitación colaborativo.

Una vez que el grupo de alumnos comienza con los ejercicios el esquema a seguir por el programa es el siguiente: se busca una pregunta del nivel correspondiente y se le presenta a los alumnos en su espacio asignado, las preguntas pueden ser de alternativas o de escritura libre. Las instrucciones asociadas a la pregunta son sintetizadas por la librería de audio y los alumnos las escuchan en sus audífonos.

Cuando todos los alumnos del grupo marcan su respuesta como lista, esta es comparada con la respuesta esperada para la pregunta, de haber alguna inconsistencia esto es reportado y se le pide a los alumnos que lo intenten de nuevo (Figura 2.2.2 – 1, primera decisión). A diferencia del sistema individual, en donde si un alumno se equivocaba 5 veces o más en la misma pregunta esta era ignorada y se procedía a la siguiente, en el sistema colaborativo no existe la opción de saltar una pregunta que el alumno no sabe porque el objetivo era buscar generar discusión entre los alumnos y se consideró que los alumnos lograrían eventualmente en conjunto llegar a la solución correcta, esto fue confirmado por la experimentación en donde en ningún caso se observó un grupo de alumnos que no lograra resolver algún problema trabajando en conjunto.

De no haber alguna inconsistencia entre las respuestas de los alumnos y la respuesta esperada las respuestas se consideran correctas. Se registra la solución exitosa de la pregunta por el grupo y luego se revisa si los alumnos están listos para subir de nivel. Un grupo se considera listo para subir de nivel cuando ha respondido en total más de un cierto número de respuestas del nivel de manera satisfactoria (10 por defecto) y las últimas n preguntas han sido respondidas correctamente en serie, siendo n un número configurable (5 por defecto) menor que la cantidad de preguntas necesarias correctas en total para el nivel (Figura 2.2.2 – 1, segunda decisión).

Cuando un grupo se considera listo para subir de nivel la siguiente pregunta que les es presentada corresponde al nuevo nivel, con contenido y estructura distinta, pero manteniendo el estándar de preguntas de alternativas o texto.

Tanto las respuestas correctas como las incorrectas incluyen feedback inmediato y personalizado para los alumnos, ya sea fomentándolos a intentar de nuevo en caso de una respuesta insatisfactoria como felicitándolos en el caso opuesto.

El nivel máximo de las preguntas disponibles a trabajar en cada sesión es un valor configurable y si un grupo llegase a completar este nivel el sistema les comienza a asignar preguntas del primer nivel, repitiendo el proceso mientras dure la sesión.

3. ARQUITECTURA DE SOFTWARE

Inicialmente, el desarrollo del sistema se realizó siguiendo la línea y estructura utilizadas por la investigación del uso de mouse para la enseñanza de la aritmética (Mouse-Based Interpersonal Computer for Individual Learning; Alcoholado et al, 2012), dado que las limitaciones parecían similares y se podía aprovechar el conocimiento y desarrollo obtenido en dicha investigación. Estos parámetros de desarrollo eran el uso del lenguaje C# y el desarrollo para el sistema operativo Windows 7. Sin embargo, la naturaleza de la aplicación implica el doble de dispositivos USB (un teclado y tarjeta de audio por alumno, en vez de un mouse por alumno) y rápidamente se hizo evidente que el sistema Windows tiene limitaciones en el manejo simultáneo de más dispositivos USB que lo normalmente empleado por una sola persona.

Se decidió probar con una versión de desarrollador del sistema operativo Windows 8, el cual en el minuto aún no se encontraba disponible para el público general. Si bien al principio se obtuvieron resultados alentadores, eventualmente se hizo evidente que las limitaciones que impedían alcanzar un número utilizable de dispositivos eran causadas por sectores del sistema operativo que no era posible alterar, debido a la naturaleza cerrada de Windows.

Frente a esta situación se decidió tomar una decisión drástica y abandonar Windows por un sistema en donde modificar las bases operativas del sistema fuese factible, para esto se eligió trabajar con Ubuntu, el sistema operativo abierto de distribución Unix más popular. Debido a la popularidad y apertura del sistema era factible modificar las bases estructurales de este y existe una gran cantidad de documentación y de comunidades de donde obtener información. Eventualmente se construyó un sistema Ubuntu con el kernel ligeramente modificado para aumentar la cantidad de dispositivos USB conectables.

Dado que el sistema Ubuntu no es capaz de ejecutar aplicaciones desarrolladas en C# nativamente y que los programas que permiten la ejecución de estas aplicaciones en este ambiente siempre tienen costos significativos en tiempo de ejecución se decidió reescribir el sistema en un lenguaje nativo para Ubuntu, y para esto se decidió utilizar Python, ya que permite desarrollo veloz y flexible de aplicaciones y existen múltiples librerías para el apoyo de los distintos contenidos de la aplicación, como PyGame, la cual fue utilizada para el desarrollo gráfico del sistema.

Para el desarrollo del sistema se decidió seguir un patrón de Modelo - Vista - Controlador, en donde el Modelo guarda la información permanente y relativa al estado del sistema y sus variables, la Vista está a cargo del manejo de la interfaz de usuario y el Controlador es el orquestador general que se encarga de la comunicación entre modelo y vista y de que el programa siga el flujo principal. Se utilizaron también fuentes externas de información, específicamente las preguntas a utilizar con los alumnos y la lista de los cursos con los que se iba a trabajar.

El siguiente es un esquema grafico de las relaciones entre las clases del sistema:

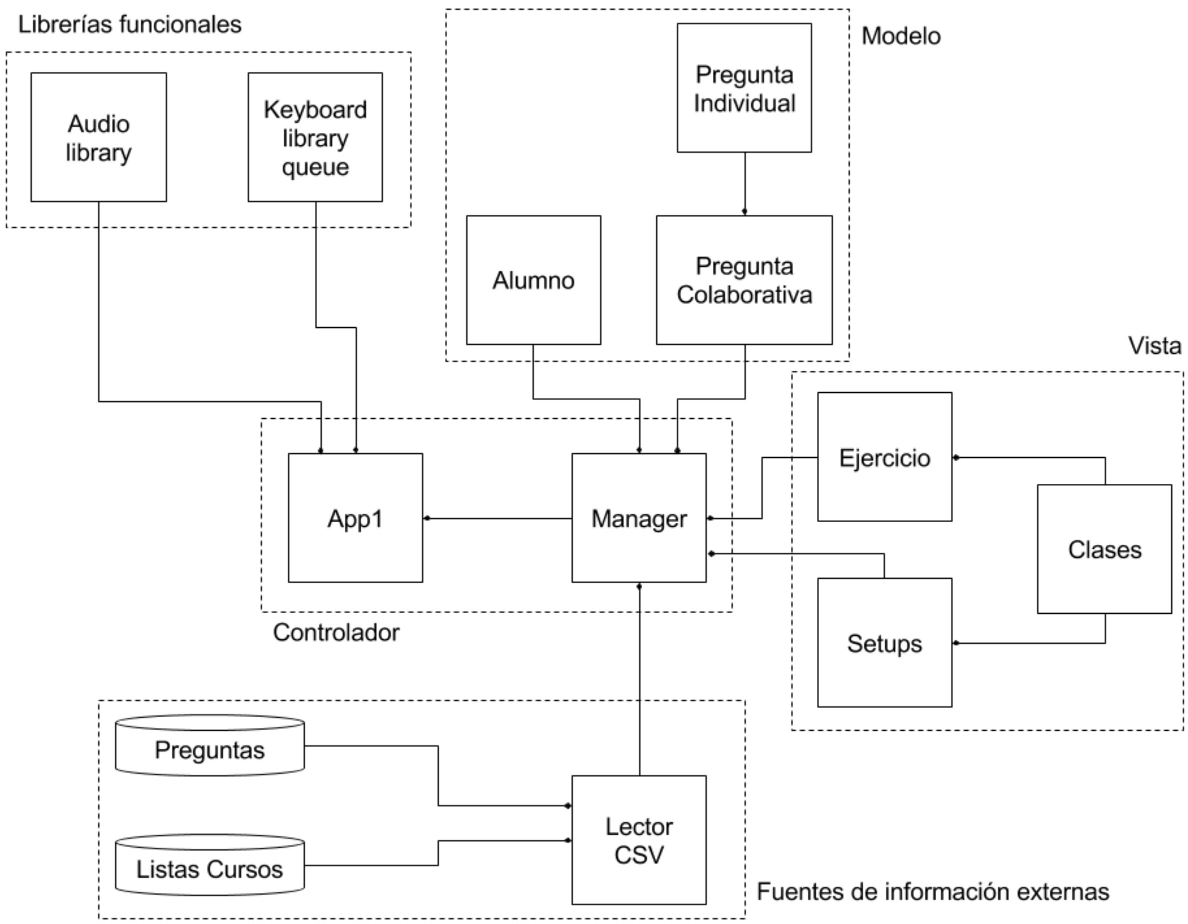

Figura 3 – 1: Esquema de clases del software.

3.1 Fuentes de información externa (Figura 3 – 1, sector inferior):

Con el fin de hacer el sistema fácilmente modificable y extensible, se decidió que las fuentes de información externa (los ejercicios y las listas de los alumnos) se almacenarán en archivos CSV que serían leídos al iniciar la aplicación. Esto permite que nuevas preguntas puedan ser agregadas de manera fácil y que solo requieran de software capaz de leer y escribir CSV, haciendo así un sistema que permite que los profesores agreguen o modifiquen las preguntas a realizarles a los alumnos sin mucho conocimiento previo de computación o programación. Otro beneficio de tener todas las preguntas en un archivo y cargarlas a memoria cuando se inicia el sistema es la baja latencia en la lectura y búsqueda de preguntas, ya que leer de memoria es órdenes de magnitud más rápido que la lectura de disco y dado que la información contenida en las preguntas consiste solamente de texto, no existe riesgo de causar problemas en la ejecución por uso excesivo de memoria.

La clase a cargo de la lectura de los archivos CSV funciona con patrón Singleton, con el propósito de limitar el uso en memoria y solo leer los archivos CSV una vez, ya que esta es una operación que puede ocupar una cantidad significativa de tiempo. La clase cuenta con un objeto estático instancia (el objeto del singleton), un método estático para obtener la instancia y un constructor de objetos, método que se encarga de efectivamente leer los archivos CSV. Los objetos de la clase tienen por su parte dos atributos, ambos diccionarios, uno que guarda la información relevante a las preguntas y otro que guarda los nombres de los alumnos. El diccionario de preguntas tiene como clave el nivel de las preguntas y como valor una lista de preguntas, las cuales son a su vez objetos de la clase PreguntaIndividual. El diccionario de alumnos, por otro lado, tiene como clave el número de lista del alumno y como valor su nombre, ya que esta es toda la información que se tiene de los alumnos en este CSV.

Los principios detrás de esta arquitectura fueron utilizados tanto para el sistema colaborativo como individual, y en algunos casos incluso el mismo código fue utilizado para ambos sistemas.

3.2 Modelo (Figura 3 – 1, sector central superior)

Las clases del tipo modelo tienen como principal responsabilidad almacenar la información del estado del sistema y saber cómo responder frente a los cambios en este. Las principales clases en esta categoría son las siguientes:

Alumno: Esta clase almacena toda la información relativa al alumno, es decir, su número de lista, nombre, teclado y audífono asignado, etc. Es gracias a esta clase que es posible mantener un registro de que niveles los alumnos completan con mayor facilidad y en qué nivel los alumnos terminan una sesión para comenzar la siguiente en ese mismo nivel.

PreguntaIndividual: Esta clase almacena la información leída de los archivos CSV de preguntas, es decir, el nivel de la pregunta, instrucciones, respuesta correcta, etc.

PreguntaColaborativa: La pregunta individual adaptada para el uso colaborativo, capaz de combinar múltiples preguntas individuales en una sola pregunta colaborativa. Eventualmente se decidió trabajar directamente con las preguntas individuales ya que los niveles que permitían combinación de preguntas manteniendo los parámetros de colaboración buscados eran muy pocos.

3.3 Vista (Figura 3 – 1, sector derecho)

Estas clases están encargadas de construir la interfaz de usuario, en base a la información recibida tanto de la interacción del usuario como del estado interno del sistema. Para generar la interfaz de usuario se utilizó la librería PyGame, librería diseñada principalmente para la construcción de juegos en Python y que otorga un gran nivel de libertad y control sobre el display, además de que permite acceso concurrente por diferentes procesos de la aplicación. Las principales clases de esta categoría son:

Ejercicio: Clase que construye el display para mostrarle al usuario, recibe una pregunta en su constructor y en base a eso construye la interfaz correcta para el tipo de pregunta. Diferentes clases dentro del mismo archivo heredan de la clase ejercicio base para poder trabajar con diversos tipos de preguntas (texto, alternativas, etc.).

Setups: Esta clase contiene la serie de vistas para mostrar al usuario al comienzo de la sesión, para el pareamiento de teclados y audífonos, la identificación del alumno y la creación de grupos. La mayor diferencia con las clases de ejercicio es que en estas vistas solo un alumno está trabajando, ya que los grupos no han sido formados todavía. Por esto son vistas de menor tamaño y complejidad.

Clases: en este archivo se encuentran una serie de clases de uso común para los ejercicios y el setup, principalmente clases desarrolladas como una utilidad para diversas partes del código. Incluye textbox para permitir input de texto, numbox para inputs numéricos (número de lista, grupo) y Listview para selección múltiple.

3.4 Controlador (Figura 3 – 1, sector central)

Las clases a cargo de controlar el flujo de la aplicación y de conectar el modelo y la vista, no mantienen información ni interactúan con el usuario, pero orquestan todo el proceso.

Manager: La clase que controla el flujo principal de los ejercicios de los alumnos, dirigiendo la información ingresada por ellos a el display para que reaccione de la manera apropiada y hacia la pregunta que los alumnos están trabajando. Existe un objeto manager por cada grupo de alumnos, el cual contiene los objetos con los que los alumnos trabajan, estos objetos (preguntas y display) van variando con el tiempo, pero el manager es el mismo durante toda la sesión.

App1: La clase principal de la aplicación, considerada dentro de los controladores porque es la clase que configura todo el sistema principal e inicia todos los procesos necesarios. Contiene todas las otras clases y es el punto de partida del sistema.

Además de estas clases que describen el funcionamiento específico del sistema, existen dos librerías desarrolladas en conjunto para la aplicación colaborativa e individual, estas son audio_library y keyboard_library_queue, para el manejo de audio y teclados respectivamente.

4. METODOLOGÍA EXPERIMENTAL:

4.1 Modelo Experimental:

Para la etapa de desarrollo de software se realizaron repetidas pruebas con alumnos universitarios para medir la estabilidad del software y el impacto en la máquina. Se realizaron repetidas sesiones a lo largo de un semestre con alumnos universitarios (Figura 4.1 – 1) en donde se encontraron múltiples puntos de falla del sistema hasta que se logró llegar a un punto de estabilidad suficiente como para ir a un colegio y trabajar con alumnos.

Figura 4.1 – 1: Alumnos universitarios con sistema en etapa experimental.

Se contactó al colegio Teniente Dagoberto Godoy y se acordó trabajar con la generación de tercero básico, que contaba de cuatro cursos mixtos de aproximadamente 40 alumnos por curso. Se decidió utilizar los cursos de la siguiente manera, un curso como grupo control, que tendrían clases regulares y controles periódicos para medir avance, un curso utilizo la versión individual del software, como herramienta de medición de la utilidad del software y los dos cursos restantes trabajaron con la versión colaborativa del software.

Para los tres cursos que trabajaron con el software (tanto la versión colaborativa como la individual) se dividió el curso en dos secciones, para que el grupo trabajando con el sistema fuese más manejable y para disminuir el stress en la máquina. Se realizaron sesiones una vez por semana con cada curso y para los cursos que trabajaron con la maquina se intercalan sesiones de evaluación (sesiones con la maquina) y sesiones de repaso.

Para medir el avance de los cursos durante el semestre, se realizó un pre y post test, una prueba escrita que ha sido utilizada para otras intervenciones y cuya validez ha sido respaldada en otros experimentos. La misma prueba se realiza antes de comenzar la intervención en los cursos y luego de terminarla y los resultados fueron comparados para determinar avance de los alumnos.

En la versión individual, los alumnos comenzaban la sesión de evaluación en el mismo nivel en el que había terminado la sesión anterior, mientras que para los grupos de tres alumnos de la versión colaborativa del software, los alumnos comenzaban la sesión en el nivel del alumno que hubiese terminado la última sesión en el nivel más bajo, es decir, si se tiene tres alumnos formando un grupo, y el último nivel de ejercicio para los alumnos del grupo registrado es nivel dos, nivel tres y nivel cuatro, el grupo comenzará por lo ejercicios del nivel dos. Esto principalmente porque se consideró más importante que ningún alumno se perdiera niveles, además de que el principio colaborativo del sistema permite que los alumnos que ya pasaron ese nivel lleven a su compañero a su propio nivel más rápido que si él estuviera trabajando solo.

Se trabajó con el colegio por un semestre, por lo que se realizaron aproximadamente 5 sesiones experimentales por curso, intercaladas con 5 sesiones de repaso.

4.2 Experimentación en Colegio

Cuando el software se considero estaba en un estado lo bastante estable y completo como para ser probado con alumnos y se había establecido un plan de acción con el colegio se procedió a comenzar las sesiones experimentales con los alumnos (Figura 4.2 – 1).

Figura 4.2 – 1: Alumnos de colegio en sesión colaborativa.

Desde la primera actividad real con los alumnos fue evidente que la realidad de la experimentación en sitio era completamente distinta a lo que se había preparado durante el desarrollo. Uno de los primeros problemas que se enfrentó en el colegio fue que la sala de clases sólo disponía de un enchufe disponible, por lo que para poder alimentar el computador, proyector y los múltiples repetidores USB se requirió de una serie de alargadores y múltiples eléctricos que aumentaron la inestabilidad del sistema, causando que en repetidas ocasiones el sistema “botara” teclados por falta de energía eléctrica en ellos para enviar señales. Una solución a este problema fue buscada con la dirección del colegio, quienes se comprometieron a instalar un segundo enchufe en la sala, aunque esto nunca ocurrió.

Para lograr una mayor continuidad en las sesiones, el sistema fue modificado de forma que si perdía un teclado por falta de señal no causará un colapso en la aplicación y el teclado fuese recuperado cuando volviera a enviar señales. También se agregó la posibilidad de eliminar un alumno de un grupo para que el resto pudiese seguir trabajando y el alumno fuese transferido a otro espacio disponible con un teclado y audífono funcionando.

Sin embargo, modificar el software para funcionar bajo estas nuevas condiciones traía sus propios problemas, ya que la cantidad de pruebas a la que se podía someter una nueva versión era limitada y por lo tanto a veces aparecían situaciones inesperadas cuando se estaba trabajando con los niños. Una forma de paliar esto era llevar dos versiones en simultáneo de la aplicación, de forma de tener la última versión "estable" disponible si es que la nueva versión tenía problemas para comenzar.

La gran mayoría de las sesiones fueron completadas exitosamente, sin embargo, se hizo evidente a lo largo del semestre que el sistema requiere de un nivel de estabilidad no disponible en la sala de clases en que se probó, por lo menos no en el estado en que se encontraba en ese minuto.

5. Resultados

Una vez terminada la fase experimental, se procedió a realizar un análisis de los datos obtenidos, estos datos provienen del pre y post test aplicado a los alumnos, los registros generados por la aplicación del comportamiento de los alumnos y por último, de las observaciones realizadas durante las sesiones de experimentación por los asistentes de investigación.

Los resultados obtenidos fueron cuantificados y analizados bajo distintos principios estadísticos, mediante el uso de software especializado, de forma de obtener indicadores estándar de análisis de pruebas, como los que siguen.

El pre test fue aplicado a 150 alumnos, mientras que el post test fue aplicado a 139, esta diferencia de 11 alumnos se explica por la ausencia de algunos alumnos a la sesión de post test y a cambios en la composición de los cursos a lo largo del semestre. Los valores obtenidos para el Alfa de Cronbach de estas pruebas fueron 0.818 y 0.775 respectivamente, los cuales están considerados dentro del rango de Bueno y/o Aceptable (George & Mallery, 2003), por lo que se sabe que se puede confiar en los valores obtenidos.

La división de los cursos en grupos experimentales se realizó de la siguiente manera:

Grupo control: Curso C

Grupo actividad individual: Curso B

Grupo actividad colaborativa: Curso A y D

Los datos estadísticos generales obtenidos para los distintos cursos están a continuación:

Grupo Control:

Estadísticos[a]

		Pre test	Post test	Diferencia	Asistencia
N	Válidos	23	23	23	23
	Perdidos	0	0	0	0
Media		26,48	30,04	3,57	91,22
Mediana		27,00	30,00	4,00	100,00
Moda		28	32	4	100
Desv. típ.		4,851	5,022	3,527	15,866
Varianza		23,534	25,225	12,439	251,723
Rango		17	19	12	67
Mínimo		19	21	-2	33
Máximo		36	40	10	100
Percentiles	25	22,00	26,00	1,00	83,00
	50	27,00	30,00	4,00	100,00
	75	29,00	32,00	7,00	100,00

a. Curso = C

Tabla 5 – 1: Resultados generales grupo control

Grupo Individual:

Estadísticos[a]

		Pre test	Post test	Diferencia	Asistencia
N	Válidos	36	36	36	35
	Perdidos	0	0	0	1
Media		28,50	29,47	,97	82,26
Mediana		28,50	30,00	1,00	83,00
Moda		23[b]	32	0	83
Desv. típ.		4,651	4,185	3,582	18,432
Varianza		21,629	17,513	12,828	339,726
Rango		18	23	15	83
Mínimo		19	15	-6	17
Máximo		37	38	9	100
Percentiles	25	25,00	27,00	-1,75	83,00
	50	28,50	30,00	1,00	83,00
	75	32,00	32,00	3,75	100,00

a. Curso = B

b. Existen varias modas. Se mostrará el menor de los valores.

Tabla 5 – 2: Resultados generales grupo individual

Grupos Colaborativos:

Estadísticos[a]

		Pre test	Post test	Diferencia	Asistencia
N	Válidos	37	37	37	37
	Perdidos	0	0	0	0
Media		24,16	27,32	3,16	92,43
Mediana		23,00	28,00	4,00	100,00
Moda		20	30	4	100
Desv. típ.		5,085	4,785	4,259	11,880
Varianza		25,862	22,892	18,140	141,141
Rango		18	22	19	40
Mínimo		15	15	-7	60
Máximo		33	37	12	100
Percentiles	25	20,50	24,00	,00	80,00
	50	23,00	28,00	4,00	100,00
	75	28,00	30,00	7,00	100,00

a. Curso = A

Tabla 5 – 3: Resultados generales grupo colaborativo A

Estadísticos[a]

		Pre test	Post test	Diferencia	Asistencia
N	Válidos	35	35	35	35
	Perdidos	0	0	0	0
Media		28,49	30,43	1,94	87,00
Mediana		29,00	30,00	2,00	100,00
Moda		27	29[b]	3	100
Desv. típ.		5,043	4,053	3,572	15,059
Varianza		25,434	16,429	12,761	226,765
Rango		18	15	15	50
Mínimo		19	23	-5	50
Máximo		37	38	10	100
Percentiles	25	25,00	28,00	-1,00	75,00
	50	29,00	30,00	2,00	100,00
	75	33,00	33,00	4,00	100,00

a. Curso = D

b. Existen varias modas. Se mostrará el menor de los valores.

Tabla 5 – 4: Resultados generales grupo colaborativo D

Al calcular valor de Cohen's d para los resultados obtenidos en los distintos cursos se obtienen los siguientes valores:

Curso	Valor	Interpretación
Control	0,72	Efecto medio
Individual	0.22	Efecto leve
Colaborativo A	0.64	Efecto medio
Colaborativo D	0.42	Efecto leve

Tabla 5 – 5: Valores de Cohen's d para los pre y post test por curso

El análisis de los datos estadísticos obtenidos no permite realizar conclusiones muy definitivas respecto al impacto de la intervención en el aprendizaje de los alumnos, ya que los resultado obtenidos en el pre test eran bastante elevados y si bien hubo aprendizaje en todos los cursos y un efecto medible, la diferencia entre el grupo control y los grupos colaborativos no es realmente significativa. Un análisis del ANOVA de un solo factor entre pre y post test refleja las siguientes correlaciones:

Grupo	Valor
Control	-0.314
Individual	-0.509
Colaborativo A	-0.487
Colaborativo D	-0.604

Tabla 5 – 6: ANOVA de un factor de diferencia pre y post test

Se observa que todos los valores son negativos, lo que significa que en todos los cursos fue el grupo que peor le fue en el pre test quien más mejoró, es decir, ambas estrategias favorecen a los alumnos que se encuentran más rezagados en el aprendizaje. Estos valores no reflejan las conclusiones obtenidas previamente con la experimentación individual sin shared display, en donde la clase tradicional favorece principalmente a los alumnos que ya iban aventajados (Sciarresi, Nussbaum 2014).

6. CONCLUSIONES

La pregunta que se planteó como motivación para este experimento es la siguiente ¿Es posible desarrollar colaboración en grupos pequeños con shared display para enseñanza del lenguaje? Existe la tentación de observar los resultados obtenidos y responder positivamente, ya que efectivamente los alumnos que trabajaron con el sistema de computador interpersonal en actividades colaborativas aprendieron y trabajaron en equipo. Sin embargo, la realidad de esta pregunta es que busca encontrar si es que hubo mejoras cualitativas o cuantitativas medibles respecto a los métodos tradicionales de enseñanza que justifiquen la inversión y tiempo dedicados al sistema. Y es frente este análisis que una respuesta es más difícil de concluir, ya que pareciera ser que al no haber una diferencia significativa entre los grupos no existe una justificación para el sistema, pero tomando en cuenta las dificultades enfrentadas durante la intervención en la sala de clases se observa otra realidad.

Si un sistema en ocasiones inestable y con comportamiento impredecible genera un nivel de aprendizaje comparable con las clases tradicionales entonces cabe considerar que de estar el sistema en un estado más estable y de haber tenido los alumnos una experiencia más controlada los resultados podrían ser más favorables. La tecnología tiene también muchas otras aplicaciones y se ha aplicado exitosamente en áreas como el aprendizaje de matemáticas, por lo que existe interés en seguir explorando los límites de estos sistemas. Es por esto que las lecciones obtenidas en esta experiencia son de fundamental importancia para la investigación que continua bajo la dirección de la alumna de doctorado, Andrea Vázquez.

Los grandes aprendizajes que se pueden sacar de esta experiencia se encuentran en el ámbito del desarrollo del software y cómo esto afectó el desempeño. Como desarrollador de software es fácil concentrarse en las abstracciones del sistema y en la complejidad del código, dejando de lado el mundo real y las máquinas en donde el software efectivamente vive. Pero el efecto que tienen las limitaciones eléctricas y las conexiones físicas fue uno de los factores más importantes en el desempeño subóptimo del sistema en el ambiente del colegio, la mayor limitación a la hora de escribir software es el ambiente real en el que este se ejecuta, y ninguna cantidad de modificaciones al kernel nos va a permitir evitar que el balanceo de cargas eléctricas en el sistema sea tan dispar que eventualmente hay hardware que simplemente no va a poder funcionar.

Estos efectos sí se pueden paliar en todo caso, sobre todo si uno planea contra ellos, y esta es la otra gran lección de la experiencia. La importancia de realizar pruebas a fondo y en diversas condiciones antes de declarar un sistema como listo para su uso real, es imposible replicar las condiciones de un colegio y de trabajar con niños sin encontrarse en esas condiciones, pero múltiples puntos de falla y condiciones poco favorables no fueron consideradas inicialmente. La mayor cantidad de los cambios introducidos en el software durante la intervención en el colegio fueron para controlar problemas que se podrían haber encontrado antes con un proceso más riguroso de evaluación de calidad.

Por último, en cuanto a diseño de software, la principal lección que se obtiene de la experiencia es que el punto más importante en este tipo de sistema es que el manejo de los recursos físicos tiene que ser lo más riguroso y flexible posible. Una de las principales mejoras que se obtuvo para la calidad de la experiencia fue cuando el sistema se volvió capaz de recuperar los teclados perdidos por desconexión, usualmente causada por falta de voltaje. Cuando se construye un sistema que tiene como una de las principales consideraciones la realidad física en donde este se va a ejecutar se está construyendo un sistema robusto que tiene cimientos fuertes.

Se puede ver cómo están lecciones han sido de gran ayuda para el futuro del proyecto y de la tecnología desarrollada, ya que las nuevas versiones del sistema han sido construidas de forma íntegra con el desarrollo físico elegido, al mismo tiempo que han sido sometidas a nuevas y continuas formas de testing de forma de tener un sistema capaz y estable antes de avanzar a una intervención en un colegio nuevamente.

En cuanto a la arquitectura desarrollada en conjunto para ambos sistemas, se puede observar que trabajar en una misma arquitectura para ambos permitió un desarrollo más veloz y dinámico, ya que ambos proyectos avanzaban en paralelo cuando se trabaja en la mayoría de las áreas. Por otra parte, desarrollar la arquitectura en abstracto de los sistemas a aplicarse permite que esta sea reutilizada en el futuro, ya que es flexible y la misma arquitectura podría ser aplicada en un sistema para enseñar otras materias o con otro tipo de contenidos.

Sin embargo, la arquitectura común de los sistemas introdujo un par de limitaciones en ellos, la primera siendo que el uso de las preguntas individuales en el sistema colaborativo limita el alcance colaborativo de estas preguntas, para los objetivos de esta intervención no fue una limitación tan grave, ya que se buscaba más aprendizaje de los contenidos que de herramientas de trabajo en grupo. Por otra parte, la arquitectura desarrollada está limitada a el estado físico de la maquina, es decir, la arquitectura desarrollada fue construida específicamente para el tipo de dispositivos utilizados en la intervención y de considerarse la posibilidad de permitir una mayor diversidad de dispositivos (e.g. distintas marcas / modelos de teclados) sería necesaria una reestructuración no menor de algunas secciones de código.

II. CAPÍTULO 2: THE IMPACT OF THE TECHNOLOGY USED IN FORMATIVE ASSESSMENT: THE CASE OF SPELLING

1. SUMMARY

Formative assessment is a set of student-centered practices, which do not always report optimum levels of learning. In addition to this, different technologies have been shown to be better suited to certain tasks than to others. We therefore look at how the technology used to assist formative assessment in spelling affects improvements in learning. This study was conducted in Chile in two phases, following the Integrative Learning Design framework and methodology. We first developed a formative assessment strategy for teaching spelling, and then studied how technology affected this strategy. This was done by comparing two different technologies: Tablet PCs and the Interpersonal Computer. We learned that a self-paced formative assessment strategy using Tablets was more effective than the same strategy using an Interpersonal Computer when teaching spelling to primary school students. We therefore showed that when using a formative assessment strategy the technology that is used has an impact on learning.

2. INTRODUCTION AND BACKGROUND

Formative assessment is a set of practices used in the teaching-learning process. These practices take into account the quality of work produced by students in order to mold and improve their skills (Sadler, 1989). Formative assessment looks to identify what the students do (and do not) know, and with this make significant changes in the learning process (Boston, 2002). This can be achieved by following a 5-step model (Black, 2015):

a. Plan the assessment and establish clear objectives

b. Develop activities that can meet these objectives

c. Carry out these activities in the classroom

d. Use informal assessment to check the degree to which the objectives have been met

e. Take this assessment into account in subsequent teaching

Formative assessment requires “moments of contingency” in teaching (steps c and d in the above model), which allow the learning process to be regulated (step e). These moments can be synchronous, such as group or class discussions, or asynchronous, such as using evidence from exercises completed by the students (Black & Wiliam, 2009).

Within the teaching-learning process, formative assessment can lead to improvements in learning. This is because it takes into consideration the students’ developing skills in order for the teacher to provide the necessary support (Gikandi, Morrow, & Davis, 2011). This is achieved by providing the teacher with a feedback loop that molds and guides student development in order to meet the learning objective (Roskos & Neuman, 2012). Despite the benefits of formative assessment, its use in the classroom is far from widespread (Shute & Kim, 2014). Part of the reason for this is that formative assessment requires teachers to learn how to obtain and analyze information from several sources regarding student learning in only a short space of time (Ruiz-Primo, 2011). This forces them to radically change the way in which they interact with their students and how they operate in the classroom (Black, 2015). In practice, this is difficult to achieve.

Literature reports differing results regarding the effects of formative assessment on student achievement in primary and secondary education (Bennett, 2011; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012). These differences might be explained by the fact that defining and implementing formative assessment in the classroom is a complex process (Antoniou & James, 2014); a process that must respect the conditions of the classroom (Rodríguez, Nussbaum, & Dombrovskaia, 2012).

One of these conditions is the use of technology. Technology is not neutral and can be better suited to some tasks than to others (Angeli & Valanides, 2009; Koehler & Mishra, 2009). It is therefore critical to understand how different technologies influence specific teaching practices (Koehler & Mishra, 2009). Within technology, Tablet PCs currently have the highest penetration rate in education (McEwen & Dubé, 2015). These devices allow for different forms of interaction and are easy to use (Zhu et al., 2014). This makes them powerful teaching devices (Milman, Carlson-Bancroft, & Boogart, 2014; Wang, Wu, Chien, Hwang, & Hsu, 2015). An alternative technology is the Interpersonal Computer (IPC). This technology allows multiple users located in the same physical space to interact simultaneously. They do so by using individual input devices (such as a mouse or keyboard), a single computer, and a shared display, such as a projector screen (Kaplan, DoLenh, & Bachour, 2009). The IPC is particularly attractive to developing countries because of its low entry cost (1 dollar per child per year) (Trucano, 2010). Furthermore, the IPC requires less technological support than a computer lab as there is only one machine and the users do not have direct access to the system's software (Alcoholado, Diaz, Tagle, Nussbaum, & Infante, 2014).

Given the disparate effects of formative assessment reported by the literature, as well as the growing use of technology to assist the learning process, it is important to study how technology affects this practice. We have chosen to use spelling as the topic in this study. This is because of the importance of teaching spelling as a linguistic skill (McNeill & Kirk, 2014), as well as the fact that improvements in learning have previously been achieved using formative assessment for developing writing skills among native speakers (Graham, Harris, & Hebert, 2011; Horstmanshof & Brownie, 2013). Given the wide variety of educational technologies that are available, it is interesting to study two very different technologies that are focused on two distinct socioeconomic groups.

Therefore, the aim of this study is to answer the research question: "How does the technology used to assist formative assessment affect improvements in student learning when studying spelling?" In order to do so, two studies were conducted in Chile within the context of teaching Spanish to native speaking students.

Based on the literature review, the following hypotheses were developed:

H1. A self-paced learning strategy using formative assessment, where the teacher is provided with information on student performance so that they can plan a targeted review class accordingly, produces significant learning gains when compared to a self-paced learning strategy with review sessions that follow the order of the curriculum.

H2. Different technological platforms have different effects on learning when following a strategy of self-paced learning using formative assessment.

3. METHODOLOGY

An Integrative Learning Design framework (ILD) (Bannan-Ritland, 2003) was used in this study to guide a process of design-based research (The Design-Based Research Collective, 2003). ILD was chosen as it was developed specifically for technological interventions in an educational setting (van den Akker, Bannan, Kelly, Nieveen, & Plomp, 2009). Based on this framework, the research was structured in two phases.

Phase one, which looked to prove the first hypothesis, focused on designing a self-paced learning strategy using formative assessment in order to improve student learning in spelling. Once the first hypothesis had been proven, the second phase was carried out. The aim of this phase was to measure the impact on learning of two different technologies for self-paced learning using formative assessment, based on the previously validated strategy.

The participating students were native Spanish speakers in the first years of primary school. Both phases of the study were conducted in government-subsidized schools in Santiago, Chile. This type of school was chosen as they have the highest percentage of student enrolment in the country (MINEDUC, 2015) and are therefore more representative of the general population.

The ILD framework highlights the importance of avoiding the practice effect, by using phase-specific “data streams” (Bannan-Ritland, 2003). Different groups of students therefore participated in each phase of the study in order to avoid this effect. The two participating schools (one for each phase) were chosen at random, as well as the classes within each school.

3.1. Variables of Study, Instruments and Data Collection

The instruments were developed in three stages. Firstly, a piece of software was designed and developed to support self-paced learning using formative assessment for teaching spelling in Spanish. Subsequently, a set of instruments was developed to measure student learning. Finally, as a result of the iterative process within the ILD framework, the need arose to incorporate another instrument in order to record student behavior during the second phase.

A piece of self-paced learning software was developed to teach spelling in Spanish. Given that students can have very different levels of knowledge (Tomlinson, 2014), the software included the entire contents of the curriculum established by the Chilean Ministry of Education for 1st to 6th grade (MINEDUC, 2012). Each topic was included as a separate level on the software, with the levels following the same sequence as they are introduced to students during their primary education.

For formative assessment to take place, the software generated a student progress report for the teacher at the end of each session. With this, the teacher could identify the main areas of difficulty for the students and use this information to prepare the following class. This class could therefore focus on the topics that the students most needed to review. Doing so provided the “moments of contingency” that are necessary for learning (Black & Wiliam, 2009), as well as defining the role of the teacher within the intervention (Urhahne, Schanze, Bell, Mansfield, & Holmes, 2010).

Once the software was developed, the following variables were defined in order to assess learning in spelling:

a. Student knowledge before the intervention (initial level)

b. Student knowledge after the intervention (final level)

c. Learning gain, i.e. the difference in student knowledge before and after the intervention

To measure these variables, a written test (pre- and post-) was used to assess the students' knowledge before and after the intervention. The test was developed based on the topics set by the Chilean Ministry of Education for teaching spelling from 1st to 6th grade (MINEDUC, 2012). Each question on the test was related to one of the levels from the software used in the study. The same test was used in both phases of this study; Cronbach's alpha for this test was 0.753.

The variables defined by Beserra, Nussbaum, Zeni, Rodriguez, & Wurman (2014) were used to assess student behavior in the second phase of the study. These variables are divided into distraction indicators and concentration indicators:

Distraction indicators:

Distraction without causing interruptions: the student stops paying attention but does not interrupt the work of others.

Interrupting classmates: the student becomes distracted and starts to interrupt their classmates' work.

Talking about an unrelated topic: the student stops working and talks with their classmates or the researchers about an unrelated topic.

Concentration indicators:

Talking about the topic of the session: the student asks for or provides help with the topic, or makes comments about the topic they are working on.

Talking about the software: the student asks for or provides help with the platform, or makes comments about the software.

Classroom observations were included in the second phase in order to measure the aforementioned variables regarding student behavior. A group of observers recorded student behavior during each of the sessions using technology during phase two. This was done by following the guidelines established by Beserra et al. (2014). These external observers were paid university students and did not interact with the participants in the study. There was one observer for every 5 students. To standardize the way in which information was gathered, the observers were trained before conducting the observations. Furthermore, the observers were also rotated so as to avoid any bias. Due to an agreement with the participating school, the review sessions were not observed.

3.2. Data Analysis

In phase one, descriptive statistics were first obtained for the results from the pre- and post-test. Following this, Levene's test and a t-test were used to assess the homogeneity of means and variance for the participating groups. These tests revealed that there was homogeneity in both groups and so an ANOVA was subsequently used to assess the difference between the groups' post-test scores. Another ANOVA was also conducted within each group to assess the difference between the pre-test and post-test scores (learning gain), as well as the respective Cohen's d in order to test the effect size. In order to deduce which students in the group benefited most from the intervention, correlations were made between pre-test scores and learning gains, as well as between post-test scores and learning gains. To assess the relationship between the number of exercises completed and learning, a correlation was also made between pre-test scores and the final level reached on the software, as well as between post-test scores and the final level reached.

In phase two, descriptive statistics were obtained for the results from the pre- and post-test. Levene's test and a t-test were also conducted to assess the homogeneity of means and variance. As the conditions were met, an ANOVA was then used for two different situations: firstly between the groups to assess the difference in post-test scores, and subsequently within each group to assess learning gains. Finally, a t-test was used to measure the difference in the number of levels completed by each of the groups. Behavioural statistics were also recorded and compared for each group in every session.

4. PHASE ONE: FORMATIVE ASSESSMENT VERSUS TRADITIONAL REVIEW CLASSES

This section describes the details of the experiment that was carried out in order to validate the first hypothesis, H1: "A self-paced learning strategy using formative assessment, where the teacher is provided with information on student performance so that they can plan a targeted review class accordingly, produces significant learning gains when compared to a self-paced learning strategy with review sessions that follow the order of the curriculum."

4.1. Context

A study was conducted in 6 sessions across 6 weeks in a government-subsidized school in Santiago de Chile. In 2015, 71% of the students enrolled in the school came from a vulnerable background (i.e. they were at risk of dropping out of school). Despite this, the school has been ranked above average for similar establishments on national standardized tests for Language Arts.

4.2. Experimental Design

The objective of this phase was to study the impact on student learning when using a technology-assisted formative assessment strategy for self-paced study of spelling.

Based on a review of the literature, we learned the following:

a. In Chile, there are huge differences between students from the same class, as the curriculum assumes that everyone learns at the same pace (Salazar, 2014). Therefore, a system to assist learning (in this context) must cover several years of the curriculum and allow the students to advance at their own pace.
b. The role of the teacher must be clearly established when incorporating technology into the teaching practice (Urhahne et al., 2010).

c. Interventions in spelling have a greater effect on student learning when these are done in primary school (Goodwin & Ahn, 2013; Puranik & Alotaiba, 2012).

All of the participating students used the software that was designed for this study. As indicated in section “Variables of Study, Instruments and Data Collection”, the software that was developed included topics for 1st to 6th grade spelling. The decision to do so was based on point (a) of the requirements obtained from the literature review for this phase. After each session using the technology, the students had a review class with their teacher. In other words, they had three sessions using the technology, interspersed with three review sessions.

Formative assessment was only used with the experimental group. In this case, the teacher used the information on the students’ performance when using the self-paced software to prepare the review class after each session using the technology. The teacher received a report on the topics that caused most difficulty for students from that group. For the control group, on the other hand, the contents of the review sessions were pre-defined at the beginning of the project. These sessions did not take into consideration the progress made by the students, nor the topics which they found most difficult when working with the self-paced software. In both cases, the review sessions were conducted by the class teacher, taking into consideration point (b) of the requirements obtained from the literature review for this phase.

In phase one, the software was implemented using Tablet PCs. Tablet PCs were chosen because they are widely used within education (McEwen & Dubé, 2015), as well as being easy to use for primary school students and teachers (Zhu et al., 2014).

4.3. Participants and Procedure

This phase featured 45 2nd grade students, aged between 7 and 8, as well as their Spanish teacher. This grade level was chosen as it is the first grade in which spelling is formally studied, while it also takes into consideration point (c) from the literature review for this phase.

The students were randomly divided into two groups: an experimental group of 18 students (8 boys, 10 girls) and a control group of 27 students (5 boys, 22 girls). Both groups were taught by the same teacher, thus controlling for the effect of the teacher on student performance (Kane, Taylor, Tyler, & Wooten, 2011; Sanders & Rivers, 1996).

A 45-minute pre-test was conducted the day before the intervention began. For the first session, the experimental group used the technology and the control group were given the first review class (with pre-defined contents). In the following sessions, the groups alternated between using the technology and having review classes. All of the sessions (review and technology) lasted 20 minutes each. A 45-minute post-test was conducted the day after the final session.

4.4. Results Phase One: Formative Assessment Versus Traditional Review Classes

	Control Group			Experimental Group		
	Pre-test total	Post-test total	Difference between pre- and post-test	Pre-test total	Post-test total	Difference between pre- and post-test
N	27			18		
Median	25.93	28.63	2.7	27.56	30.61	3.05
Std. Dev.	4.632	5.644	5.090	5.338	4.146	3.654
Range	17	23	19	19	13	14
Minimum	20	19	-4	19	24	-6
Maximum	37	42	15	38	37	8

Table 4.4 - 1: Descriptive statistics for pre- and post-test in phase one

Table 4.4 - 1 shows the descriptive statistics for the pre- and post-test. Levene's test shows that there is homogeneity of variance between the two groups ($F = 1.328$, $p = .256$). On average, the experimental group scored a point higher on the pre-test than the control group. However, this difference is not statistically significant ($t = 1.057$, $p = .298$) and therefore the two groups are comparable. With this, the conditions for conducting an analysis of variance (ANOVA) are therefore satisfied.

The ANOVA for the difference in post-test results between the experimental group and control group reveal that there is no evidence to suggest that this difference is significant, $F(1,43) = 0.064$, $p = .802$.

The results of an ANOVA for the difference between the pre- and post-test results for reveals the near-significant increases in learning for both groups, control, $F(1,42) = 3.927$, $p = .053$; and experimental, $F(1,34) = 3.676$, $p = .064$.

Quintile	Sum of the squares	Root mean square	F	Sig.
1	0.4091	0.4091	1.385	0.36
2	0.5	0,5	3	0.333
3	6	3	0	1
4	0.6944	0.6944	25	0.03777
5	56.25	56.25	225	0.00442

Table 4.4 - 2: ANOVA for the difference in scores, separated by quintiles, for phase one.

To analyze this result in greater depth, Table 4.4 - 2 reveals the results from the ANOVA for the difference between the pre- and post-test results (learning gain). These are broken down into quintiles of students for the experimental and control groups. Each quintile contained 9 students. The first quintile corresponds to the students with the smallest difference between their pre- and post-test scores, while the fifth quintile corresponds to the students with the greatest difference. This difference is significant for the final two quintiles. On average, the experimental group showed greater improvements in learning (Table 4.4 - 1). The results of the ANOVA by quintile therefore suggest that students with the greatest improvements in learning benefitted significantly more from targeted review (Cohen's d = 0.517 for quintile 4, Cohen's d = 0.61 for quintile 5).

A Pearson product-moment correlation analysis reveals a positive correlation between the post-test score and the pre-post difference in the control group (r = .632, p = .0001). This correlation is not present in the experimental group (r = .068, p = .790). This suggests that the students from the control group who performed better on the post-test managed a greater improvement in learning by the end of the study. Therefore, students in the control group with a higher level of knowledge of the topics covered in the study benefited more from the traditional review method.

Finally, there is a positive correlation between the number of levels completed on the software and the post-test score for both experimental ($r = .543$, $p = .022$) and control groups ($r = .510$, $p = .007$). This correlation is not present when considering the number of levels completed on the software and the pre-test score (experimental group: $r = .318$, $p = .199$; control group: $r = .284$, $p = .151$). This means that in both groups the students who completed more levels on the software scored higher on the post-test. Given that there is no correlation with pre-test scores, this suggests that the skills practice led to improved performance on the post-test.

4.5. Discussion and Limitations of Phase One Experiment

The results from this phase suggest that a self-paced learning strategy using formative assessment and assisted by technology leads to significant learning gains for students who are given target review session based on their specific needs. Significant learning gains are not achieved when using the same technology and traditional review sessions, i.e. where the contents follow the order of the curriculum without taking into account the students' specific needs.

One of the main limitations of this phase is the scope of the results. There are two factors that could help more students experience significant learning gains. Firstly, the technology component (the self-paced learning software) was only implemented using Tablets. There are certain platforms that are better suited than others to different tasks

(Koehler & Mishra, 2009) and it is therefore necessary to explore the effect on learning when using other devices for formative assessment. Secondly, it is important to take into consideration that students within a classroom have different levels of skill development (Tomlinson, 2014). This should be expressed in their work with the self-paced learning software.

5. PHASE TWO: FORMATIVE ASSESSMENT SUPPORTED BY TABLETS VERSUS FORMATIVE ASSESSMENT SUPPORTED BY INTERPERSONAL COMPUTER

This section describes the details of the experiment that was conducted in order to validate the second hypothesis, H2: “Different technological platforms have different effects on learning when following a strategy of self-paced learning using formative assessment.”

5.1. Context

This second phase was also conducted in 6 sessions across 6 weeks in a government-subsidized school in Santiago de Chile. This school was not the same one that participated in phase one, although their socioeconomic characteristics mean they are comparable. In 2015, 75% of the students enrolled in the school came from a vulnerable background (i.e. they were in danger of dropping out of school). This school has achieved above-average scores on national standardized tests for Language Arts when compared to similar establishments.

5.2. Experimental Design

Given the results from phase one, the objective of this second literature review was to find evidence to ensure an improvement in learning for all students. Our findings are summarized as follows:

a. Different levels of improvement in learning are reported when using formative assessment in primary education (Bennett, 2011; Briggs et al., 2012; Coffey, Hammer, Levin, & Grant, 2011). Furthermore, we see that technology is better suited to some tasks than to others (Angeli & Valanides, 2009; Koehler & Mishra, 2009). This may affect learning when implementing different teaching strategies.
b. Placements tests can be used to gauge the students' initial level of knowledge (Nielson, 2011; Yang, Chuang, Li, & Tseng, 2013). For technological interventions using self-paced learning, this allows the more advanced students to make better use of the available time by starting at a higher level, in line with their prior knowledge.

Considering this, the objective of the second phase was to measure the impact on learning of two different technologies using the same formative assessment strategy to teach spelling.

Given point (a) of the requirements obtained from the literature review for this phase, the experience was performed with two groups; one with Tablet PCs and the other with IPC (adapting the software to this last technology). The progress reports generated after each session by both technologies were used to prepare the review sessions for both groups.

Furthermore, given point (b) of the requirements obtained from the literature review for this phase, the pre-test was used as a placement test (Nielson, 2011; Yang et al., 2013) so that the students started working from a level that was in line with their prior knowledge. In this sense, the students did not necessarily start from the most basic level and were therefore able to make better use of the time that was available.

This phase followed the same procedure as that described for the experimental group in phase one. The differences were that each student was placed at a different starting level depending on their performance on the pre-test, as well as the use of two different technologies to support the self-paced learning software.

5.3. Participants and Procedure

The study featured 46 students from two 3rd grade classes, aged between 8 and 9. 3rd grade students were used in this phase as the participating school did not formally teach spelling until this grade.

The children were randomly divided into two groups (Tablet and IPC). The groups included students from both classes so as to control for this factor. Both groups had 23 students. The Tablet group featured 13 girls and 10 boys, while the IPC group featured 8 girls and 15 boys. As with phase one, both groups of students were taught by the same teacher in order to control for the effect of the teacher on learning (Kane et al., 2011; Sanders & Rivers, 1996).

Students from both groups, Tablet and IPC, had three sessions using technology, interspersed with three review sessions based on formative assessment.

As with phase one, the students sat a 45-minute pre-test one day before the intervention began. Both groups had three sessions with the technology, interspersed with three targeted review sessions based on the results of each child's self-paced learning. A 45-minute post-test was sat by both groups the day after the final session.

5.4. Results Phase Two: Formative Assessment Supported by Tablets Versus Formative Assessment Supported by Interpersonal Computer

	Tablet Group		**IPC Group**	
	Pre-test	**Post-test**	**Pre-test**	**Post-test**
N (students)	23		23	
Min	11	15	13	13
Max	24	26	25	26
Average	17.652	20.130	19.391	19.478
St. Dev.	3.432	2.833	3.407	2.936

Table 5.4 - 1: Descriptive statistics for both groups in phase two.

Based on the data from Table 5.4 - 1, Levene's test reveals that there is homogeneity of variance between the groups ($F = 0.033$, $p = .856$). This table also shows that the groups started with a similar score on the pre-test ($t = 1.724$, $p = .09$) and therefore makes them comparable. It is therefore possible to study these results using an ANOVA.

By the end of the study, the difference in post-test scores between the Tablet and IPC groups was not significant ($t = 0.860$, $p = .394$). However, the difference in average scores between the pre- and post-test by students from the Tablet group was statistically significant ($t = 2.046$, $p = .047$, Cohen's $d = 0.603$). This was not the case for the students from the IPC group ($t = 0.843$, $p = .403$).

If we include all of the students who participated in 2 or more of the sessions (regardless of whether or not they took the pre- and/or post-test), there is a significant difference between the number of levels completed for both groups ($t = 2.447$, $p = .01$). In this case, the number of levels completed is taken as the final level reached by a student, minus their starting level. This shows that the group using Tablets managed to make greater progress.

With regards to the observations made during the technology sessions, the students in the Tablet group participated more actively than the IPC students. This is demonstrated by the greater number of positive events recorded by the observers (263.8 observations per session for the Tablet group versus 114 for the IPC group). It is worth remembering that the students were randomly assigned to each group. It is therefore possible to have groups with similar characteristics, which means that their behavior when using the different platforms can be compared.

	Session 1		Session 2		Session 3	
	Tablet	**IPC**	**Tablet**	**IPC**	**Tablet**	**IPC**
Interrupting	0.47	0.11	0.06	0.03	0.40	0.00
Distraction without causing interruptions	2.63	0.66	1.56	0.23	0.12	0.07
Talking about another topic	0.68	0.05	0.53	0.00	0.64	0.00
Talking about the topic of the session	3.79	0.79	3.39	0.54	4.12	1.13
Talking about the software	2.74	0.39	1.75	0.72	0.76	2.13

Table 5.4 - 2: Evolution of student behavior (number of observations per student) for each session, for both groups in phase two.

Table 5.4 - 2 reveals the evolution in student behavior for each session. The data is standardized per student (the ratio between the number of observations recorded for each variable and the number of students present). It shows that in both groups students rarely interrupted one another or talked about unrelated topics. However, this is slightly lower in the case of the IPC group. Distractions without interrupting other students decrease across the sessions in both groups, and are greater among the Tablet group. Talking about the topic of the session remains relatively constant for both groups, although this is always greater for the Tablet group. However, talking about the software behaves inversely as it starts higher for the Tablet group and lower for the IPC group, with both finishing the other way round.

5.5. Discussion and Limitations of Phase Two Experiment

In phase two, significant improvements in learning were observed among students using Tablet PCs. However, this was not the case for students using the Interpersonal Computer. It was also observed that students from the Tablet group made greater progress on the software by completing more levels than the students from the Interpersonal Computer group. By doing so, students from the Tablet group therefore covered more of the curriculum. It must be noted that the initial level of knowledge was comparable for the two groups (see section “Results Phase Two”), and that both groups received the same sort of targeted review sessions. It can therefore be inferred that the platform that was used had an effect on student progress with the software and, therefore, on the amount of topics that they covered.

The classroom observations conducted during this phase showed that students from the Tablet group participated more actively than students from the IPC group. They also revealed that students from the Tablet group started to talk less about the software and more about the topic of the session as the sessions went on. The opposite was true for the IPC group. Given that the students were randomly assigned to each of the platforms, the Tablet and IPC groups should, on average, comprise students with similar behavioral characteristics. This leads us to conclude that the platforms have properties that foster certain behavior among students. We can therefore conclude from phase two that there are certain features of Tablets that favor learning in this context.

Within the limitations of this phase, it is worth noting that Tablets are aimed at individual work, while the IPC is aimed more at cooperative/collaborative work as it uses a shared screen (Nussbaum, Alcoholado, & Büchi, 2015). The impact of the devices on learning may therefore be related to the type of learning strategy that is employed. For this study, the strategy was based on individual work and, therefore, Tablets may have been better suited to the teaching objective. It remains as future work to study the effect of the IPC (or other shared screen technologies) on similar strategies that feature group work.

6. SYNTHESIS AND DISCUSSION

The following summary can be made of the main findings in order to assess what has been learnt in these two phases of the study, in terms of the relationship between formative assessment and technology:

a. A self-paced learning strategy using formative assessment fosters learning among the students that are targeted by the formative component: the data from both phases reveals the importance of the contents of the review class, while the second phase revealed the importance of starting the exercises on the software from a level that is in line with the student’s prior knowledge. This is consistent with the findings by Tomlinson (2014) regarding the need to differentiate the class and contents in order to adapt to a heterogeneous classroom.

b. Certain characteristics of technological platforms make them better suited to supporting a particular learning strategy: given the formative assessment strategy that was designed for this study, in this particular case it was observed that Tablets fostered significant learning gains. This may be due to the fact that the children’s interaction with the Tablets, unlike the IPC, is very similar to a note pad and may therefore aid the process of teaching spelling (Neumann & Neumann, 2013).

In terms of the study's initial objectives, Hypothesis 1 was proven to be valid as there were differences in learning when employing a self-paced learning strategy using formative assessment, as compared to using the same self-paced learning strategy without formative assessment (targeted review). Furthermore, Hypothesis 2 showed that Tablets produced a significant difference in learning for the type of formative assessment strategy designed for this study.

In summary, and in response to the research question "How does the technology used to assist formative assessment affect improvements in student learning when studying spelling?", we can conclude that two elements must be considered when designing a formative assessment strategy supported by technology. Firstly, the formative component (review class and exercises) must respond to the students' specific needs. Secondly, the strategy must be supported by a technological platform that fosters student behaviour that is in line with the objective of the teaching strategy.

7. CONCLUSIONS

Formative assessment is a set of student-centered practices, which do not always report optimum levels of learning. Furthermore, different technologies have been shown to be better suited to certain tasks than to others. We therefore look at how the technology used to assist formative assessment in spelling affects improvements in learning.

The first phase of the study showed us that self-paced learning using formative assessment with a teacher providing targeted review improved learning among lower-performing students. This could be explained by the fact that formative assessment identifies what the students do (and do not) know and focuses the teacher's work on the students' specific needs (Luckin et al., 2012; OECD & CERI, 2008). Students with greater knowledge of the topic benefited more from the traditional review sessions in the group that used the self-paced learning software without formative assessment. This could be explained by the fact that the teacher delivered the content at a pace that could be followed by the more advanced students, without focusing on the difficulties faced by the other students.

The more moderate levels of learning observed among the higher-performing students in the group with formative assessment from the first phase of the study was addressed by using a placement test. This test allows the students to start working from a level that is more in line with their prior knowledge (Nielson, 2011; Yang et al., 2013). Doing so therefore allows the students to make better use of the time that is available. This component was added to the self-paced learning software in the second phase of the study, while maintaining the formative assessment strategy employed in this study.

The aim of the second phase was to answer our research question: "How does the technology used to assist formative assessment affect improvements in student learning when studying spelling?". The results from this phase revealed that when following the same strategy, students using self-paced learning with Tablets experienced significant learning gains, which was not the case for students using the Interpersonal Computer (IPC). This may be due to the fact that Tablets, unlike the IPC, foster student behavior that is in line with the objective of the teaching strategy used.

We therefore showed that when using a formative assessment strategy, the technology that is used has an impact on learning.

Based on the results, a series of guidelines can be developed for designing formative assessment strategies. When working on spelling with native speakers in primary school, it is recommended teaching in two differentiated stages in order to improve learning. The first stage must feature self-paced exercises, starting from a level that is consistent with each student's level of prior knowledge. The second stage refers to review classes, which must target the students' specific needs and address the heterogeneity of the classroom (Tomlinson, 2014). It is important to address the needs of lower-performing students, while not forgetting their higher-performing counterparts. Doing so allows learning to be fostered among all of the students.

Technology can help with the implementation of formative assessment strategies. It is therefore important to bear in mind the relationship between the two. The platform that is chosen must be in line with the teaching objectives, as certain technologies are better suited than others depending on the type of work (e.g. cooperative/collaborative or individual).

One of the limitations of this study is its scope. Both phases of the study (phase one and two) were based on the same topic (Spanish spelling for children aged between 7 and 9, native Spanish speakers); lasted for 6 weeks; and consisted of 3 sessions using technology, interspersed with 3 review sessions. It remains as future work to conduct a more extensive study covering different topics and a wider range of ages.

Our results also have to be validated by other formative assessment practices. For example, using strategies that foster collaborative work or that consider formative methods other than lecture-based classes.

Finally, the way in which other technologies, such as smartphones or laptops, affect our results should also be studied.

III. BIBLIOGRAFÍA

Alcoholado, C., Nussbaum, M., Tagle, A., Gomez, F., Denardin, F., Susaeta, H., & Toyama, K.. (2012). One Mouse per Child: interpersonal computer for individual arithmetica practice. Journal of Computer Assisted Learning, 28(4), 295-309.

Alcoholado, C., Diaz, A., Tagle, A., Nussbaum, M., & Infante, C. (2014). Comparing the use of the interpersonal computer, personal computer and pen-and-paper when solving arithmetic exercises. British Journal of Educational Technology, n/a–n/a. http://doi.org/10.1111/bjet.12216

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). Computers & Education, 52(1), 154–168. http://doi.org/10.1016/j.compedu.2008.07.006

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. Educational Assessment, Evaluation and Accountability, 26(2), 153–176. http://doi.org/10.1007/s11092-013-9188-4

Bannan-Ritland, B. (2003). The role of design in research: The Integrative Learning Design framework. Educational Researcher, 32(1), 21–24. http://doi.org/10.3102/0013189X032001021

Bax, S. (2003). CALL—past, present and future. System.

Bennett, R. E. (2011). Formative assessment: A critical review. Assessment in Education: Principles, Policy & Practice, 18(1), 5–25. http://doi.org/10.1080/0969594X.2010.513678

Beserra, V., Nussbaum, M., Zeni, R., Rodriguez, W., & Wurman, G. (2014). Practising arithmetic using educational video games with an Interpersonal Computer. Educational Technology & Society, 17(3), 343–358. Retrieved from http://www.ifets.info/journals/17_3/26.pdf

Black, P. (2015). Formative assessment – an optimistic but incomplete vision. Assessment in Education: Principles, Policy & Practice, 22(1), 161–177. http://doi.org/10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31. http://doi.org/10.1007/s11092-008-9068-5

Boston, C. (2002). The concept of formative assessment. Practical Assessment, Research & Evaluation, 8(9). Retrieved from http://eric.ed.gov/?id=ED470206

Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. Educational Measurement: Issues and Practice, 31(4), 13–17. http://doi.org/10.1111/j.1745-3992.2012.00251.x

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. Journal of Research in Science Teaching, 48(10), 1109–1136. http://doi.org/10.1002/tea.20440

Danesh, A., Inkpen, K.M., Lau, F., Shu, K., & Booth, K.S. (2001). GeneyTM: Designing a collaborative activity for the palm Handheld computer. Proceedings of the Conference on Human Factors in Computing Systems (CHI 2001), Seattle, USA, 388-395.

Darling-Hammond L. (2000). Teacher quality and student achievement: a review of state policy and evidence. Education Policy Analysis Archives.

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. Computers & Education, 57(4), 2333–2351. http://doi.org/10.1016/j.compedu.2011.06.004

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in english: Effects on literacy outcomes for school-age children. Scientific Studies of Reading, 17(4), 257–285. http://doi.org/10.1080/10888438.2012.689791

Graham, S., Harris, K., & Hebert, M. (2011). Informing Writing: The Benefits of Formative Assessment. A Carnegie Corporation Time to Act report. Washington, DC. Retrieved from http://carnegie.org/fileadmin/Media/Publications/InformingWriting.pdf

Horstmanshof, L., & Brownie, S. (2013). A scaffolded approach to discussion board use for formative assessment of academic writing skills. Assessment & Evaluation in Higher Education, 38(1), 61–73. http://doi.org/10.1080/02602938.2011.604121

Jipping, M., Dieter, S., Krikke, J., & Sandro, S. (2001). Using Handheld Computers in the Classroom: Laboratories and Collaboration on Handheld Machines, paper with, Proceedings of the 2001 SIGCSE Technical Symposium, SIGCSE Technical Bulletin. 33, 1 Charlotte, North Carolina, United States, 169-173.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. Journal of Human Resources, 46(3), 587–613. http://doi.org/10.1353/jhr.2011.0010

Kaplan, F., DoLenh, S., & Bachour, K. (2009). Interpersonal computers for higher education. Interactive Artifacts and Furniture Supporting Collaborative Work and Learning, 10, 1–17. http://doi.org/10.1007/978-0-387-77234-9_8

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? Contemporary Issues in Technology and Teacher Education, 9(1), 60–70. Retrieved from http://www.editlib.org/p/29544/

Luckin, R., Bligh, B., Manches, A., Ainsworth, S., Crook, C., & Noss, R. (2012). Decoding learning: Promise and digital education. London. Retrieved from http://www.nesta.org.uk/publications/decoding-learning

McEwen, R., & Dubé, A. K. (2015). Intuitive or idiomatic: An interdisciplinary study of child-tablet computer interaction. Journal of the Association for Information Science and Technology. http://doi.org/10.1002/asi.23470

McNeill, B., & Kirk, C. (2014). Theoretical beliefs and instructional practices used for teaching spelling in elementary classrooms. Reading and Writing, 27(3), 535–554. http://doi.org/10.1007/s11145-013-9457-0

Milman, N. B., Carlson-Bancroft, A., & Boogart, A. (2014). Examining differentiation and utilization of iPads across content areas in an independent, PreK–4th grade elementary school. Computers in the Schools, 31(3), 119–133. http://doi.org/10.1080/07380569.2014.931776

MINEDUC. (2012). Programa de estudio de lenguaje y comunicación. Santiago. Retrieved from http://curriculumenlinea.mineduc.cl/sphider/search.php?query=&t_busca=1&results=&sesear=1&dis=0&category=10#a6921

MINEDUC. (2015). Variación de matrícula y tasas de permanencia por sector. Santiago, Chile. Retrieved from http://centroestudios.mineduc.cl/tp_enlaces/portales/tp5996f8b7cm96/uploadImg/File/Evidencias/Evidencias final_julio_2015.pdf

Neumann, M. M., & Neumann, D. L. (2013). Touch screen tablets and emergent literacy. Early Childhood Education Journal, 42(4), 231–239. http://doi.org/10.1007/s10643-013-0608-3

Nielson, K. (2011). Self-study with language learning software in the workplace: What happens. Language Learning & Technology, 15(3), 110–129. Retrieved from http://www.llt.msu.edu/issues/october2011/v15n3.pdf#page=115

Nussbaum, M., Alcoholado, C., & Büchi, T. (2015). A comparative analysis of interactive arithmetic learning in the classroom and computer lab. Computers in Human Behavior, 43, 183–188. http://doi.org/10.1016/j.chb.2014.10.031

Organisation for Economic Co-operation and Development, & Centre for Educational Research and Innovation. (2008). Assessment for learning. Formative assessment. “Learning in the 21st Century: Research, Innovation and Policy.” Retrieved from http://www2.glos.ac.uk/offload/tli/lets/lathe/issue1/articles/brown.pdf

Prabhu, N. S. (1987). Second language pedagogy. Oxford: Oxford University Press.

Puranik, C. S., & Alotaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. Reading and Writing, 25(7), 1523–1546. http://doi.org/10.1007/s11145-011-9331-x

Rodríguez, P., Nussbaum, M., & Dombrovskaia, L. (2012). Evolutionary development: A model for the design, implementation, and evaluation of ICT for education programmes. Journal of Computer Assisted Learning, 28(2), 81–98. http://doi.org/10.1111/j.1365-2729.2011.00419.x

Roskos, K., & Neuman, S. B. (2012). Formative assessment: Simply, no additives. The Reading Teacher, 65(8), 534–538. http://doi.org/10.1002/TRTR.01079

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), Computer-Supported Collaborative Learning, 69-97. New York: Springer-Verlag.

Roggof, B. (1994). Developing understanding of the idea of communities of learners. Mind, Culture and Activity, 1(4), 209-229.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students’ learning. Studies in Educational Evaluation, 37(1), 15–24. http://doi.org/10.1016/j.stueduc.2011.04.003

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18(2), 119–144. Retrieved from http://www.jstor.org/stable/info/23369143

Salazar, P. (2014). Casi dos tercios de alumnos de 8° básico sabe menos que uno de 5° en matemáticas. La Tercera. Santiago. Retrieved from http://www.latercera.com/noticia/nacional/2014/11/680-602737-9-casi-dos-tercios-de-alumnos-de-8-basico-sabe-menos-que-uno-de-5-en-matematicas.shtml

Salomon, G. (1990). The Computer Lab: A Bad Idea Now Sanctified. Educational Technology, 30(10), 50-52.

Sanders, W., & Rivers, J. (1996). Cumulative and residual effects of teachers on future student academic achievement. Tenesse. Retrieved from http://mccluelearning.com/wp-content/uploads/2011/09/Cumulative-and-Residual-Effects-of-Teachers.pdf

Sciarresi, E., Nussbaum, M. (2014). Evaluación del impacto en el aprendizaje de la instrucción de ortografía, considerando el conocimiento de los alumnos de la materia de estudio.

Scott, S.D., Shoemaker, G.B.D., and Inkpen, K.M. (2000). Towards seamless support of natural collaborative interactions. Proceedings of Graphics Interface. May, Montreal, Canada, 103-110.

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), Handbook of Research on Educational Communications and Technology (Vol. 13, pp. 311–321). Springer New York. http://doi.org/10.1007/978-1-4614-3185-5

Sotomayor, C., Molina, D., Bedwell, P., & Hernández, C. (2013). Caracterización de problemas ortográficos recurrentes en alumnos de escuelas municipales chilenas de 3º, 5º y 7º básico. Signos. Estudios de Lingüística, 105-131.

Stanton, D., Neale, H., & Bayon, V. (2002). Interfaces to Support Children's Co-present Collaboration: Multiple Mice and Tangible Technologies. In Gerry Stahl, editor, Computer Support for Collaborative Learning (CSCL), Lawrence Erlbaum Associates, 342-352.

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. Educational Researcher, 32(1), 5–8. Retrieved from http://www.jstor.org/stable/3699927

Tomlinson, C. A. (2014). Differentiated Classroom: Responding to the Needs of all Learners (2nd ed.). ASCD. Retrieved from http://books.google.com/books?hl=en&lr=&id=CLigAwAAQBAJ&pgis=1

Trucano, M. (2010). One mouse per child. Retrieved April 7, 2015, from http://blogs.worldbank.org/edutech/one-mouse-per-child

UNESCO. (2010). Escritura. Un estudio de las habilidades de los estudiantes de América Latina y el Caribe. Santiago.

Urhahne, D., Schanze, S., Bell, T., Mansfield, A., & Holmes, J. (2010). Role of the teacher in computer-supported collaborative inquiry learning. International Journal of Science Education, 32(2), 221–243. http://doi.org/10.1080/09500690802516967

Van den Akker, J., Bannan, B., Kelly, A. E., Nieveen, N., & Plomp, T. (2009). Introduction to educational design research. In T. Plomp & N. Nieveen (Eds.), Seminar conducted at the East China Normal University, Shanghai (PR china). Retrieved from http://www.slo.nl/downloads/2009/Introduction_20to_20education_20design_20research.pdf/

Vygotsky, L. S. (1979). El Desarrollo de las Funciones Psicológicas Superiores. Barcelona, Ed. Crítica.

Wang, J.-Y., Wu, H.-K., Chien, S.-P., Hwang, F.-K., & Hsu, Y.-S. (2015). Designing applications for physics learning: Facilitating high school students' conceptual understanding by using tablet PCs. Journal of Educational Computing Research, 51(4), 441–458. http://doi.org/10.2190/EC.51.4.d

Yang, Y.-T. C., Chuang, Y.-C., Li, L.-Y., & Tseng, S.-S. (2013). A blended learning environment for individualized English listening and speaking integrating critical thinking. Computers & Education, 63, 285–305. http://doi.org/10.1016/j.compedu.2012.12.012

Zhu, S., Shi, Y., Wu, D., Yang, H. H., Wang, J., & Kwok, L. (2014). To be or not to be: Using tablet PCs in K-12 education. In 2014 International Conference of Educational Innovation through Technology (pp. 220–224). IEEE. http://doi.org/10.1109/EITT.2014.42

IV. APENDICES

1. **CARTA DE RECEPCIÓN**

Manuscript ID JCAL-15-256 - Journal of Computer Assisted Learning

---

jcal.editor@gmail.com <jcal.editor@gmail.com> 9 October 2015 at 13:47

To: afvasque@uc.cl
Cc: afvasque@uc.cl, mn@ing.puc.cl, efsciarr@uc.cl, t.martinezg@gmail.com, cebaraho1@gmail.com, kstrass@uc.cl

09-Oct-2015

Dear Dr. Vasquez:

Your revised manuscript entitled "The Impact of the Technology Used in Formative Assessment: the Case of Spelling" by Vasquez, Andrea; Nussbaum, Miguel; Sciarresi, Enzo; Martinez, Tomas; Barahona, Camila; Strasser, Katherine, has been successfully submitted online.

If you are reading this as a co-authors: please contact the Editorial Office as soon as possible if you disagree with being listed as a co-author for this manuscript.

Your manuscript ID is JCAL-15-256.

Please mention the above manuscript ID in all future correspondence. If there are any changes in your postal address or e-mail address, please log in to ScholarOne Manuscripts at https://mc.manuscriptcentral.com/jcal and edit your user information as appropriate.

You can also view the status of your manuscript at any time by checking your Author Centre after logging in to https://mc.manuscriptcentral.com/jcal.

Thank you for submitting your manuscript to the Journal of Computer Assisted Learning.

Liesbeth Kester
Associate Editor, Journal of Computer Assisted Learning

---

2. ARTÍCULO ENVIADO A REVISTA (CAPÍTULO 2 EN FORMATO REVISTA)

The Impact of the Technology Used in Formative Assessment: the Case of Spelling

Abstract

Formative assessment is a set of student-centered practices, which do not always report optimum levels of learning. In addition to this, different technologies have been shown to be better suited to certain tasks than to others. We therefore look at how the technology used to assist formative assessment in spelling affects improvements in learning. This study was conducted in Chile in two phases, following the Integrative Learning Design framework and methodology. We first developed a formative assessment strategy for teaching spelling, and then studied how technology affected this strategy. This was done by comparing two different technologies: Tablet PCs and the Interpersonal Computer. We learned that a self-paced formative assessment strategy using Tablets was more effective than the same strategy using an Interpersonal Computer when teaching spelling to primary school students. We therefore showed that when using a formative assessment strategy the technology that is used has an impact on learning.

Keywords: Formative assessment; Interactive learning environments; Self-paced learning; Design-based research; Spelling; Spanish language

Formative assessment is a set of practices used in the teaching-learning process. These practices take into account the quality of work produced by students in order to mold and improve their skills (Sadler, 1989). Formative assessment looks to identify what the students do (and do not) know, and with this make significant changes in the learning process (Boston, 2002). This can be achieved by following a 5-step model (Black, 2015):

a. Plan the assessment and establish clear objectives

b. Develop activities that can meet these objectives

c. Carry out these activities in the classroom

d. Use informal assessment to check the degree to which the objectives have been met

e. Take this assessment into account in subsequent teaching

Formative assessment requires “moments of contingency” in teaching (steps c and d in the above model), which allow the learning process to be regulated (step e). These moments can be synchronous, such as group or class discussions, or asynchronous, such as using evidence from exercises completed by the students (Black & Wiliam, 2009).

Within the teaching-learning process, formative assessment can lead to improvements in learning. This is because it takes into consideration the students’ developing skills in order for the teacher to provide the necessary support (Gikandi, Morrow, & Davis, 2011). This is achieved by providing the teacher with a feedback loop that molds and guides student development in order to meet the learning objective

(Roskos & Neuman, 2012). Despite the benefits of formative assessment, its use in the classroom is far from widespread (Shute & Kim, 2014). Part of the reason for this is that formative assessment requires teachers to learn how to obtain and analyze information from several sources regarding student learning in only a short space of time (Ruiz-Primo, 2011). This forces them to radically change the way in which they interact with their students and how they operate in the classroom (Black, 2015). In practice, this is difficult to achieve.

Literature reports differing results regarding the effects of formative assessment on student achievement in primary and secondary education (Bennett, 2011; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012). These differences might be explained by the fact that defining and implementing formative assessment in the classroom is a complex process (Antoniou & James, 2014); a process that must respect the conditions of the classroom (Rodríguez, Nussbaum, & Dombrovskaia, 2012).

One of these conditions is the use of technology. Technology is not neutral and can be better suited to some tasks than to others (Angeli & Valanides, 2009; Koehler & Mishra, 2009). It is therefore critical to understand how different technologies influence specific teaching practices (Koehler & Mishra, 2009). Within technology, Tablet PCs currently have the highest penetration rate in education (McEwen & Dubé, 2015). These devices allow for different forms of interaction and are easy to use (Zhu et al., 2014). This makes them powerful teaching devices (Milman, Carlson-Bancroft, & Boogart, 2014; Wang, Wu, Chien, Hwang, & Hsu, 2015). An alternative technology is the Interpersonal Computer (IPC). This technology allows multiple users located in the same physical space to interact simultaneously. They do so by using individual input devices (such as a mouse or keyboard), a single computer, and a shared display, such as a projector screen (Kaplan, DoLenh, & Bachour, 2009). The IPC is particularly attractive to developing countries because of its low entry cost (1 dollar per child per year) (Trucano, 2010). Furthermore, the IPC requires less technological support than a computer lab as there is only one machine and the users do not have direct access to the system's software (Alcoholado, Diaz, Tagle, Nussbaum, & Infante, 2014).

Given the disparate effects of formative assessment reported by the literature, as well as the growing use of technology to assist the learning process, it is important to study how technology affects this practice. We have chosen to use spelling as the topic in this study. This is because of the importance of teaching spelling as a linguistic skill (McNeill & Kirk, 2014), as well as the fact that improvements in learning have previously been achieved using formative assessment for developing writing skills among native speakers (Graham, Harris, & Hebert, 2011; Horstmanshof & Brownie, 2013). Given the wide variety of educational technologies that are available, it is interesting to study two very different technologies that are focused on two distinct socioeconomic groups.

Therefore, the aim of this study is to answer the research question: "How does the technology used to assist formative assessment affect improvements in student learning when studying spelling?" In order to do so, two studies were conducted in Chile within the context of teaching Spanish to native speaking students.

Based on the literature review, the following hypotheses were developed:

H1. A self-paced learning strategy using formative assessment, where the teacher is provided with information on student performance so that they can plan a targeted review class accordingly, produces significant learning gains when compared to a self-paced learning strategy with review sessions that follow the order of the curriculum.

H2. Different technological platforms have different effects on learning when following a strategy of self-paced learning using formative assessment.

Methodology

An Integrative Learning Design framework (ILD) (Bannan-Ritland, 2003) was used in this study to guide a process of design-based research (The Design-Based Research Collective, 2003). ILD was chosen as it was developed specifically for technological interventions in an educational setting (van den Akker, Bannan, Kelly, Nieveen, & Plomp, 2009). Based on this framework, the research was structured in two phases.

Phase one, which looked to prove the first hypothesis, focused on designing a self-paced learning strategy using formative assessment in order to improve student learning in spelling. Once the first hypothesis had been proven, the second phase was carried out. The aim of this phase was to measure the impact on learning of two different technologies for self-paced learning using formative assessment, based on the previously validated strategy.

The participating students were native Spanish speakers in the first years of primary school. Both phases of the study were conducted in government-subsidized schools in Santiago, Chile. This type of school was chosen as they have the highest percentage of student enrolment in the country (MINEDUC, 2015) and are therefore more representative of the general population.

The ILD framework highlights the importance of avoiding the practice effect, by using phase-specific “data streams” (Bannan-Ritland, 2003). Different groups of students therefore participated in each phase of the study in order to avoid this effect. The two participating schools (one for each phase) were chosen at random, as well as the classes within each school.

Variables of Study, Instruments and Data Collection

The instruments were developed in three stages. Firstly, a piece of software was designed and developed to support self-paced learning using formative assessment for teaching spelling in Spanish. Subsequently, a set of instruments was developed to measure student learning. Finally, as a result of the iterative process within the ILD framework, the need arose to incorporate another instrument in order to record student behavior during the second phase.

A piece of self-paced learning software was developed to teach spelling in Spanish. Given that students can have very different levels of knowledge (Tomlinson, 2014), the software included the entire contents of the curriculum established by the Chilean Ministry of Education for 1st to 6th grade (MINEDUC, 2012). Each topic was included as a separate level on the software, with the levels following the same sequence as they are introduced to students during their primary education.

For formative assessment to take place, the software generated a student progress report for the teacher at the end of each session. With this, the teacher could identify the main areas of difficulty for the students and use this information to prepare the following class. This class could therefore focus on the topics that the students most needed to review. Doing so provided the “moments of contingency” that are necessary for learning

(Black & Wiliam, 2009), as well as defining the role of the teacher within the intervention (Urhahne, Schanze, Bell, Mansfield, & Holmes, 2010).

Once the software was developed, the following variables were defined in order to assess learning in spelling:

a. Student knowledge before the intervention (initial level)

b. Student knowledge after the intervention (final level)

c. Learning gain, i.e. the difference in student knowledge before and after the intervention

To measure these variables, a written test (pre- and post-) was used to assess the students' knowledge before and after the intervention. The test was developed based on the topics set by the Chilean Ministry of Education for teaching spelling from 1st to 6th grade (MINEDUC, 2012). Each question on the test was related to one of the levels from the software used in the study. The same test was used in both phases of this study; Cronbach's alpha for this test was 0.753.

The variables defined by Beserra, Nussbaum, Zeni, Rodriguez, & Wurman (2014) were used to assess student behavior in the second phase of the study. These variables are divided into distraction indicators and concentration indicators:

a. Distraction indicators:
   - Distraction without causing interruptions: the student stops paying attention but does not interrupt the work of others.
   - Interrupting classmates: the student becomes distracted and starts to interrupt their classmates' work.

- Talking about an unrelated topic: the student stops working and talks with their classmates or the researchers about an unrelated topic.

b. Concentration indicators:

- Talking about the topic of the session: the student asks for or provides help with the topic, or makes comments about the topic they are working on.
- Talking about the software: the student asks for or provides help with the platform, or makes comments about the software.

Classroom observations were included in the second phase in order to measure the aforementioned variables regarding student behavior. A group of observers recorded student behavior during each of the sessions using technology during phase two. This was done by following the guidelines established by Beserra et al. (2014). These external observers were paid university students and did not interact with the participants in the study. There was one observer for every 5 students. To standardize the way in which information was gathered, the observers were trained before conducting the observations. Furthermore, the observers were also rotated so as to avoid any bias. Due to an agreement with the participating school, the review sessions were not observed.

Data Analysis

In phase one, descriptive statistics were first obtained for the results from the pre- and post-test. Following this, Levene’s test and a t-test were used to assess the homogeneity of means and variance for the participating groups. These tests revealed that there was homogeneity in both groups and so an ANOVA was subsequently used to assess the difference between the groups' post-test scores. Another ANOVA was also conducted within each group to assess the difference between the pre-test and post-test scores (learning gain), as well as the respective Cohen's d in order to test the effect size. In order to deduce which students in the group benefited most from the intervention, correlations were made between pre-test scores and learning gains, as well as between post-test scores and learning gains. To assess the relationship between the number of exercises completed and learning, a correlation was also made between pre-test scores and the final level reached on the software, as well as between post-test scores and the final level reached.

In phase two, descriptive statistics were obtained for the results from the pre- and post-test. Levene's test and a t-test were also conducted to assess the homogeneity of means and variance. As the conditions were met, an ANOVA was then used for two different situations: firstly between the groups to assess the difference in post-test scores, and subsequently within each group to assess learning gains. Finally, a t-test was used to measure the difference in the number of levels completed by each of the groups. Behavioural statistics were also recorded and compared for each group in every session.

Phase One: Formative Assessment Versus Traditional Review Classes

This section describes the details of the experiment that was carried out in order to validate the first hypothesis, H1: "A self-paced learning strategy using formative assessment, where the teacher is provided with information on student performance so that they can plan a targeted review class accordingly, produces significant learning gains when compared to a self-paced learning strategy with review sessions that follow the order of the curriculum."

Context

A study was conducted in 6 sessions across 6 weeks in a government-subsidized school in Santiago de Chile. In 2015, 71% of the students enrolled in the school came from a vulnerable background (i.e. they were at risk of dropping out of school). Despite this, the school has been ranked above average for similar establishments on national standardized tests for Language Arts.

Experimental Design

The objective of this phase was to study the impact on student learning when using a technology-assisted formative assessment strategy for self-paced study of spelling.

Based on a review of the literature, we learned the following:

a. In Chile, there are huge differences between students from the same class, as the curriculum assumes that everyone learns at the same pace (Salazar, 2014). Therefore, a system to assist learning (in this context) must cover several years of the curriculum and allow the students to advance at their own pace.
b. The role of the teacher must be clearly established when incorporating technology into the teaching practice (Urhahne et al., 2010).
c. Interventions in spelling have a greater effect on student learning when these are done in primary school (Goodwin & Ahn, 2013; Puranik & Alotaiba, 2012).

All of the participating students used the software that was designed for this study. As indicated in section "Variables of Study, Instruments and Data Collection", the software that was developed included topics for 1st to 6th grade spelling. The decision to do so was based on point (a) of the requirements obtained from the literature review for this phase. After each session using the technology, the students had a review class with their teacher. In other words, they had three sessions using the technology, interspersed with three review sessions.

Formative assessment was only used with the experimental group. In this case, the teacher used the information on the students' performance when using the self-paced software to prepare the review class after each session using the technology. The teacher received a report on the topics that caused most difficulty for students from that group. For the control group, on the other hand, the contents of the review sessions were pre-defined at the beginning of the project. These sessions did not take into consideration the progress made by the students, nor the topics which they found most difficult when working with the self-paced software. In both cases, the review sessions were conducted by the class teacher, taking into consideration point (b) of the requirements obtained from the literature review for this phase.

In phase one, the software was implemented using Tablet PCs. Tablet PCs were chosen because they are widely used within education (McEwen & Dubé, 2015), as well as being easy to use for primary school students and teachers (Zhu et al., 2014).

Participants and Procedure

This phase featured 45 2nd grade students, aged between 7 and 8, as well as their Spanish teacher. This grade level was chosen as it is the first grade in which spelling is formally studied, while it also takes into consideration point (c) from the literature review for this phase.

The students were randomly divided into two groups: an experimental group of 18 students (8 boys, 10 girls) and a control group of 27 students (5 boys, 22 girls). Both groups were taught by the same teacher, thus controlling for the effect of the teacher on student performance (Kane, Taylor, Tyler, & Wooten, 2011; Sanders & Rivers, 1996).

A 45-minute pre-test was conducted the day before the intervention began. For the first session, the experimental group used the technology and the control group were given the first review class (with pre-defined contents). In the following sessions, the groups alternated between using the technology and having review classes. All of the sessions (review and technology) lasted 20 minutes each. A 45-minute post-test was conducted the day after the final session.

Results Phase One: Formative Assessment Versus Traditional Review Classes

(Table 1 goes here)

Table 1 shows the descriptive statistics for the pre- and post-test. Levene's test shows that there is homogeneity of variance between the two groups ($F = 1.328$, $p = .256$). On average, the experimental group scored a point higher on the pre-test than the control group. However, this difference is not statistically significant ($t = 1.057$, $p = .298$) and therefore the two groups are comparable. With this, the conditions for conducting an analysis of variance (ANOVA) are therefore satisfied.

The ANOVA for the difference in post-test results between the experimental group and control group reveal that there is no evidence to suggest that this difference is significant, $F(1,43) = 0.064$, $p = .802$.

The results of an ANOVA for the difference between the pre- and post-test results for reveals the near-significant increases in learning for both groups, control, $F(1,42) = 3.927$, $p = .053$; and experimental, $F(1,34) = 3.676$, $p = .064$.

(Table 2 goes here)

To analyze this result in greater depth, Table 2 reveals the results from the ANOVA for the difference between the pre- and post-test results (learning gain). These are broken down into quintiles of students for the experimental and control groups. Each quintile contained 9 students. The first quintile corresponds to the students with the smallest difference between their pre- and post-test scores, while the fifth quintile corresponds to the students with the greatest difference. This difference is significant for the final two quintiles. On average, the experimental group showed greater improvements in learning (Table 1). The results of the ANOVA by quintile therefore suggest that students with the greatest improvements in learning benefitted significantly more from targeted review (Cohen's $d = 0.517$ for quintile 4, Cohen's $d = 0.61$ for quintile 5).

A Pearson product-moment correlation analysis reveals a positive correlation between the post-test score and the pre-post difference in the control group ($r = .632$, $p = .0001$). This correlation is not present in the experimental group ($r = .068$, $p = .790$). This suggests that the students from the control group who performed better on the post-test managed a greater improvement in learning by the end of the study. Therefore, students in the control group with a higher level of knowledge of the topics covered in the study benefited more from the traditional review method.

Finally, there is a positive correlation between the number of levels completed on the software and the post-test score for both experimental (r = .543, p = .022) and control groups (r = .510, p = .007). This correlation is not present when considering the number of levels completed on the software and the pre-test score (experimental group: r = .318, p = .199; control group: r = .284, p = .151). This means that in both groups the students who completed more levels on the software scored higher on the post-test. Given that there is no correlation with pre-test scores, this suggests that the skills practice led to improved performance on the post-test.

Discussion and Limitations of Phase One Experiment

The results from this phase suggest that a self-paced learning strategy using formative assessment and assisted by technology leads to significant learning gains for students who are given target review session based on their specific needs. Significant learning gains are not achieved when using the same technology and traditional review sessions, i.e. where the contents follow the order of the curriculum without taking into account the students' specific needs.

One of the main limitations of this phase is the scope of the results. There are two factors that could help more students experience significant learning gains. Firstly, the technology component (the self-paced learning software) was only implemented using Tablets. There are certain platforms that are better suited than others to different tasks (Koehler & Mishra, 2009) and it is therefore necessary to explore the effect on learning when using other devices for formative assessment. Secondly, it is important to take into consideration that students within a classroom have different levels of skill development (Tomlinson, 2014). This should be expressed in their work with the self-paced learning software.

Phase Two: Formative Assessment Supported by Tablets Versus Formative Assessment Supported by Interpersonal Computer

This section describes the details of the experiment that was conducted in order to validate the second hypothesis, H2: “Different technological platforms have different effects on learning when following a strategy of self-paced learning using formative assessment.”

Context

This second phase was also conducted in 6 sessions across 6 weeks in a government-subsidized school in Santiago de Chile. This school was not the same one that participated in phase one, although their socioeconomic characteristics mean they are comparable. In 2015, 75% of the students enrolled in the school came from a vulnerable background (i.e. they were in danger of dropping out of school). This school has achieved above-average scores on national standardized tests for Language Arts when compared to similar establishments.

Experimental Design

Given the results from phase one, the objective of this second literature review was to find evidence to ensure an improvement in learning for all students. Our findings are summarized as follows:

a. Different levels of improvement in learning are reported when using formative assessment in primary education (Bennett, 2011; Briggs et al., 2012; Coffey, Hammer, Levin, & Grant, 2011). Furthermore, we see that technology is better suited to some tasks than to others (Angeli & Valanides, 2009; Koehler & Mishra, 2009). This may affect learning when implementing different teaching strategies.

b. Placements tests can be used to gauge the students' initial level of knowledge (Nielson, 2011; Yang, Chuang, Li, & Tseng, 2013). For technological interventions using self-paced learning, this allows the more advanced students to make better use of the available time by starting at a higher level, in line with their prior knowledge.

Considering this, the objective of the second phase was to measure the impact on learning of two different technologies using the same formative assessment strategy to teach spelling.

Given point (a) of the requirements obtained from the literature review for this phase, the experience was performed with two groups; one with Tablet PCs and the other with IPC (adapting the software to this last technology). The progress reports generated after each session by both technologies were used to prepare the review sessions for both groups.

Furthermore, given point (b) of the requirements obtained from the literature review for this phase, the pre-test was used as a placement test (Nielson, 2011; Yang et al., 2013) so that the students started working from a level that was in line with their prior knowledge. In this sense, the students did not necessarily start from the most basic level and were therefore able to make better use of the time that was available.

This phase followed the same procedure as that described for the experimental group in phase one. The differences were that each student was placed at a different starting level depending on their performance on the pre-test, as well as the use of two different technologies to support the self-paced learning software.

Participants and Procedure

The study featured 46 students from two 3rd grade classes, aged between 8 and 9. 3rd grade students were used in this phase as the participating school did not formally teach spelling until this grade.

The children were randomly divided into two groups (Tablet and IPC). The groups included students from both classes so as to control for this factor. Both groups had 23 students. The Tablet group featured 13 girls and 10 boys, while the IPC group featured 8 girls and 15 boys. As with phase one, both groups of students were taught by the same teacher in order to control for the effect of the teacher on learning (Kane et al., 2011; Sanders & Rivers, 1996).

Students from both groups, Tablet and IPC, had three sessions using technology, interspersed with three review sessions based on formative assessment.

As with phase one, the students sat a 45-minute pre-test one day before the intervention began. Both groups had three sessions with the technology, interspersed with three targeted review sessions based on the results of each child's self-paced learning. A 45-minute post-test was sat by both groups the day after the final session.

Results Phase Two: Formative Assessment Supported by Tablets Versus Formative Assessment Supported by Interpersonal Computer

(Table 3 goes here)

Based on the data from Table 3, Levene's test reveals that there is homogeneity of variance between the groups (F = 0.033, p = .856). This table also shows that the groups started with a similar score on the pre-test (t = 1.724, p = .09) and therefore makes them comparable. It is therefore possible to study these results using an ANOVA.

By the end of the study, the difference in post-test scores between the Tablet and IPC groups was not significant (t = 0.860, p =.394). However, the difference in average scores between the pre- and post-test by students from the Tablet group was statistically significant (t = 2.046, p = .047, Cohen's d = 0.603). This was not the case for the students from the IPC group (t = 0.843, p = .403).

If we include all of the students who participated in 2 or more of the sessions (regardless of whether or not they took the pre- and/or post-test), there is a significant difference between the number of levels completed for both groups (t = 2.447, p = .01). In this case, the number of levels completed is taken as the final level reached by a student, minus their starting level. This shows that the group using Tablets managed to make greater progress.

With regards to the observations made during the technology sessions, the students in the Tablet group participated more actively than the IPC students. This is demonstrated by the greater number of positive events recorded by the observers (263.8 observations per session for the Tablet group versus 114 for the IPC group). It is worth remembering that the students were randomly assigned to each group. It is therefore possible to have groups with similar characteristics, which means that their behavior when using the different platforms can be compared.

(Table 4 goes here)

Table 4 reveals the evolution in student behavior for each session. The data is standardized per student (the ratio between the number of observations recorded for each variable and the number of students present). It shows that in both groups students rarely interrupted one another or talked about unrelated topics. However, this is slightly lower in the case of the IPC group. Distractions without interrupting other students decrease across the sessions in both groups, and are greater among the Tablet group. Talking about the topic of the session remains relatively constant for both groups, although this is always greater for the Tablet group. However, talking about the software behaves inversely as it starts higher for the Tablet group and lower for the IPC group, with both finishing the other way round.

Discussion and Limitations of Phase Two Experiment

In phase two, significant improvements in learning were observed among students using Tablet PCs. However, this was not the case for students using the Interpersonal Computer. It was also observed that students from the Tablet group made greater progress on the software by completing more levels than the students from the Interpersonal Computer group. By doing so, students from the Tablet group therefore covered more of the curriculum. It must be noted that the initial level of knowledge was comparable for the two groups (see section “Results Phase Two”), and that both groups received the same sort of targeted review sessions. It can therefore be inferred that the platform that was used had an effect on student progress with the software and, therefore, on the amount of topics that they covered.

The classroom observations conducted during this phase showed that students from the Tablet group participated more actively than students from the IPC group. They also revealed that students from the Tablet group started to talk less about the software and more about the topic of the session as the sessions went on. The opposite was true for the IPC group. Given that the students were randomly assigned to each of the platforms, the Tablet and IPC groups should, on average, comprise students with similar behavioral characteristics. This leads us to conclude that the platforms have properties that foster certain behavior among students. We can therefore conclude from phase two that there are certain features of Tablets that favor learning in this context.

Within the limitations of this phase, it is worth noting that Tablets are aimed at individual work, while the IPC is aimed more at cooperative/collaborative work as it uses a shared screen (Nussbaum, Alcoholado, & Büchi, 2015). The impact of the devices on learning may therefore be related to the type of learning strategy that is employed. For this study, the strategy was based on individual work and, therefore, Tablets may have been better suited to the teaching objective. It remains as future work to study the effect of the IPC (or other shared screen technologies) on similar strategies that feature group work.

Synthesis and Discussion

The following summary can be made of the main findings in order to assess what has been learnt in these two phases of the study, in terms of the relationship between formative assessment and technology:

a. A self-paced learning strategy using formative assessment fosters learning among the students that are targeted by the formative component: the data from both phases reveals the importance of the contents of the review class, while the second phase revealed the importance of starting the exercises on the software from a level that is in line with the student's prior knowledge. This is consistent with the findings by Tomlinson (2014) regarding the need to differentiate the class and contents in order to adapt to a heterogeneous classroom.

b. Certain characteristics of technological platforms make them better suited to supporting a particular learning strategy: given the formative assessment strategy that was designed for this study, in this particular case it was observed that Tablets fostered significant learning gains. This may be due to the fact that the children's interaction with the Tablets, unlike the IPC, is very similar to a note pad and may therefore aid the process of teaching spelling (Neumann & Neumann, 2013).

In terms of the study's initial objectives, Hypothesis 1 was proven to be valid as there were differences in learning when employing a self-paced learning strategy using formative assessment, as compared to using the same self-paced learning strategy without formative assessment (targeted review). Furthermore, Hypothesis 2 showed that Tablets produced a significant difference in learning for the type of formative assessment strategy designed for this study.

In summary, and in response to the research question "How does the technology used to assist formative assessment affect improvements in student learning when studying spelling?", we can conclude that two elements must be considered when designing a formative assessment strategy supported by technology. Firstly, the formative component (review class and exercises) must respond to the students' specific needs. Secondly, the strategy must be supported by a technological platform that fosters student behaviour that is in line with the objective of the teaching strategy.

Conclusions

Formative assessment is a set of student-centered practices, which do not always report optimum levels of learning. Furthermore, different technologies have been shown to be better suited to certain tasks than to others. We therefore look at how the technology used to assist formative assessment in spelling affects improvements in learning.

The first phase of the study showed us that self-paced learning using formative assessment with a teacher providing targeted review improved learning among lower-performing students. This could be explained by the fact that formative assessment identifies what the students do (and do not) know and focuses the teacher's work on the students' specific needs (Luckin et al., 2012; OECD & CERI, 2008). Students with greater knowledge of the topic benefited more from the traditional review sessions in the group that used the self-paced learning software without formative assessment. This could be explained by the fact that the teacher delivered the content at a pace that could be followed by the more advanced students, without focusing on the difficulties faced by the other students.

The more moderate levels of learning observed among the higher-performing students in the group with formative assessment from the first phase of the study was addressed by using a placement test. This test allows the students to start working from a level that is more in line with their prior knowledge (Nielson, 2011; Yang et al., 2013). Doing so therefore allows the students to make better use of the time that is available. This component was added to the self-paced learning software in the second phase of the study, while maintaining the formative assessment strategy employed in this study.

The aim of the second phase was to answer our research question: “How does the technology used to assist formative assessment affect improvements in student learning when studying spelling?”. The results from this phase revealed that when following the same strategy, students using self-paced learning with Tablets experienced significant learning gains, which was not the case for students using the Interpersonal Computer (IPC). This may be due to the fact that Tablets, unlike the IPC, foster student behavior that is in line with the objective of the teaching strategy used.

We therefore showed that when using a formative assessment strategy, the technology that is used has an impact on learning.

Based on the results, a series of guidelines can be developed for designing formative assessment strategies. When working on spelling with native speakers in primary school, it is recommended teaching in two differentiated stages in order to improve learning. The first stage must feature self-paced exercises, starting from a level that is consistent with each student’s level of prior knowledge. The second stage refers to review classes, which must target the students’ specific needs and address the heterogeneity of the classroom (Tomlinson, 2014). It is important to address the needs of lower-performing students, while not forgetting their higher-performing counterparts. Doing so allows learning to be fostered among all of the students.

Technology can help with the implementation of formative assessment strategies. It is therefore important to bear in mind the relationship between the two. The platform that is chosen must be in line with the teaching objectives, as certain technologies are better suited than others depending on the type of work (e.g. cooperative/collaborative or individual).

One of the limitations of this study is its scope. Both phases of the study (phase one and two) were based on the same topic (Spanish spelling for children aged between 7 and 9, native Spanish speakers); lasted for 6 weeks; and consisted of 3 sessions using technology, interspersed with 3 review sessions. It remains as future work to conduct a more extensive study covering different topics and a wider range of ages.

Our results also have to be validated by other formative assessment practices. For example, using strategies that foster collaborative work or that consider formative methods other than lecture-based classes.

Finally, the way in which other technologies, such as smartphones or laptops, affect our results should also be studied.

REFERENCES

Alcoholado, C., Diaz, A., Tagle, A., Nussbaum, M., & Infante, C. (2014). Comparing the use of the interpersonal computer, personal computer and pen-and-paper when solving arithmetic exercises. *British Journal of Educational Technology*, n/a–n/a. http://doi.org/10.1111/bjet.12216

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, *52*(1), 154–168. http://doi.org/10.1016/j.compedu.2008.07.006

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability*, *26*(2), 153–176. http://doi.org/10.1007/s11092-013-9188-4

Bannan-Ritland, B. (2003). The role of design in research: The Integrative Learning Design framework. *Educational Researcher*, *32*(1), 21–24. http://doi.org/10.3102/0013189X032001021

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. http://doi.org/10.1080/0969594X.2010.513678

Beserra, V., Nussbaum, M., Zeni, R., Rodriguez, W., & Wurman, G. (2014). Practising arithmetic using educational video games with an Interpersonal Computer. *Educational Technology & Society*, *17*(3), 343–358. Retrieved from http://www.ifets.info/journals/17_3/26.pdf

Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 161–177. http://doi.org/10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. http://doi.org/10.1007/s11092-008-9068-5

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation*, *8*(9). Retrieved from http://eric.ed.gov/?id=ED470206

Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, *31*(4), 13–17. http://doi.org/10.1111/j.1745-3992.2012.00251.x

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, *48*(10), 1109–1136. http://doi.org/10.1002/tea.20440

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, *57*(4), 2333–2351. http://doi.org/10.1016/j.compedu.2011.06.004

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in english: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*(4), 257–285. http://doi.org/10.1080/10888438.2012.689791

Graham, S., Harris, K., & Hebert, M. (2011). *Informing Writing: The Benefits of Formative Assessment. A Carnegie Corporation Time to Act report.* Washington, DC. Retrieved from http://carnegie.org/fileadmin/Media/Publications/InformingWriting.pdf

Horstmanshof, L., & Brownie, S. (2013). A scaffolded approach to discussion board use for formative assessment of academic writing skills. *Assessment & Evaluation in Higher Education*, *38*(1), 61–73. http://doi.org/10.1080/02602938.2011.604121

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, *46*(3), 587–613. http://doi.org/10.1353/jhr.2011.0010

Kaplan, F., DoLenh, S., & Bachour, K. (2009). Interpersonal computers for higher education. *Interactive Artifacts and Furniture Supporting Collaborative Work and Learning*, *10*, 1–17. http://doi.org/10.1007/978-0-387-77234-9_8

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70. Retrieved from http://www.editlib.org/p/29544/

Luckin, R., Bligh, B., Manches, A., Ainsworth, S., Crook, C., & Noss, R. (2012). *Decoding learning: Promise and digital education*. London. Retrieved from http://www.nesta.org.uk/publications/decoding-learning

McEwen, R., & Dubé, A. K. (2015). Intuitive or idiomatic: An interdisciplinary study of child-tablet computer interaction. *Journal of the Association for Information Science and Technology*. http://doi.org/10.1002/asi.23470

McNeill, B., & Kirk, C. (2014). Theoretical beliefs and instructional practices used for teaching spelling in elementary classrooms. *Reading and Writing*, *27*(3), 535–554. http://doi.org/10.1007/s11145-013-9457-0

Milman, N. B., Carlson-Bancroft, A., & Boogart, A. (2014). Examining differentiation and utilization of iPads across content areas in an independent, PreK–4th grade elementary school. *Computers in the Schools*, *31*(3), 119–133. http://doi.org/10.1080/07380569.2014.931776

MINEDUC. (2012). *Programa de estudio de lenguaje y comunicación*. Santiago. Retrieved from http://curriculumenlinea.mineduc.cl/sphider/search.php?query=&t_busca=1&results=&sesear=1&dis=0&category=10#a6921

MINEDUC. (2015). *Variación de matrícula y tasas de permanencia por sector*. Santiago, Chile. Retrieved from http://centroestudios.mineduc.cl/tp_enlaces/portales/tp5996f8b7cm96/uploadImg/File/Evidencias/Evidencias final_julio_2015.pdf

Neumann, M. M., & Neumann, D. L. (2013). Touch screen tablets and emergent literacy. *Early Childhood Education Journal*, *42*(4), 231–239. http://doi.org/10.1007/s10643-013-0608-3

Nielson, K. (2011). Self-study with language learning software in the workplace: What happens. *Language Learning & Technology*, *15*(3), 110–129. Retrieved from http://www.llt.msu.edu/issues/october2011/v15n3.pdf#page=115

Nussbaum, M., Alcoholado, C., & Büchi, T. (2015). A comparative analysis of interactive arithmetic learning in the classroom and computer lab. *Computers in Human Behavior*, *43*, 183–188. http://doi.org/10.1016/j.chb.2014.10.031

Organisation for Economic Co-operation and Development, & Centre for Educational Research and Innovation. (2008). *Assessment for learning. Formative assessment*. *“Learning in the 21st Century: Research, Innovation and Policy.”* Retrieved from http://www2.glos.ac.uk/offload/tli/lets/lathe/issue1/articles/brown.pdf

Puranik, C. S., & Alotaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. *Reading and Writing*, *25*(7), 1523–1546. http://doi.org/10.1007/s11145-011-9331-x

Rodríguez, P., Nussbaum, M., & Dombrovskaia, L. (2012). Evolutionary development: A model for the design, implementation, and evaluation of ICT for education programmes. *Journal of Computer Assisted Learning*, *28*(2), 81–98. http://doi.org/10.1111/j.1365-2729.2011.00419.x

Roskos, K., & Neuman, S. B. (2012). Formative assessment: Simply, no additives. *The Reading Teacher*, *65*(8), 534–538. http://doi.org/10.1002/TRTR.01079

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students’ learning. *Studies in Educational Evaluation*, *37*(1), 15–24. http://doi.org/10.1016/j.stueduc.2011.04.003

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. Retrieved from http://www.jstor.org/stable/info/23369143

Salazar, P. (2014). Casi dos tercios de alumnos de 8° básico sabe menos que uno de 5° en matemáticas. *La Tercera*. Santiago. Retrieved from http://www.latercera.com/noticia/nacional/2014/11/680-602737-9-casi-dos-tercios-de-alumnos-de-8-basico-sabe-menos-que-uno-de-5-en-matematicas.shtml

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Tenesse. Retrieved from http://mccluelearning.com/wp-content/uploads/2011/09/Cumulative-and-Residual-Effects-of-Teachers.pdf

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (Vol. 13, pp. 311–321). Springer New York. http://doi.org/10.1007/978-1-4614-3185-5

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8. Retrieved from http://www.jstor.org/stable/3699927

Tomlinson, C. A. (2014). *Differentiated Classroom: Responding to the Needs of all Learners* (2nd ed.). ASCD. Retrieved from http://books.google.com/books?hl=en&lr=&id=CLigAwAAQBAJ&pgis=1

Trucano, M. (2010). One mouse per child. Retrieved April 7, 2015, from http://blogs.worldbank.org/edutech/one-mouse-per-child

Urhahne, D., Schanze, S., Bell, T., Mansfield, A., & Holmes, J. (2010). Role of the teacher in computer-supported collaborative inquiry learning. *International Journal of Science Education*, *32*(2), 221–243. http://doi.org/10.1080/09500690802516967

Van den Akker, J., Bannan, B., Kelly, A. E., Nieveen, N., & Plomp, T. (2009). Introduction to educational design research. In T. Plomp & N. Nieveen (Eds.), *Seminar conducted at the East China Normal University, Shanghai (PR china)*. Retrieved from http://www.slo.nl/downloads/2009/Introduction_20to_20education_20design_20research.pdf/

Wang, J.-Y., Wu, H.-K., Chien, S.-P., Hwang, F.-K., & Hsu, Y.-S. (2015). Designing applications for physics learning: Facilitating high school students' conceptual understanding by using tablet PCs. *Journal of Educational Computing Research*, *51*(4), 441–458. http://doi.org/10.2190/EC.51.4.d

Yang, Y.-T. C., Chuang, Y.-C., Li, L.-Y., & Tseng, S.-S. (2013). A blended learning environment for individualized English listening and speaking integrating critical thinking. *Computers & Education*, *63*, 285–305. http://doi.org/10.1016/j.compedu.2012.12.012

Zhu, S., Shi, Y., Wu, D., Yang, H. H., Wang, J., & Kwok, L. (2014). To be or not to be: Using tablet PCs in K-12 education. In *2014 International Conference of Educational Innovation through Technology* (pp. 220–224). IEEE. http://doi.org/10.1109/EITT.2014.42

Table 1 Descriptive statistics for pre- and post-test in phase one

	Control Group			**Experimental Group**		
	Pre-test total	**Post-test total**	**Difference between pre- and post-test**	**Pre-test total**	**Post-test total**	**Difference between pre- and post-test**
N	27			18		
Median	25.93	28.63	2.7	27.56	30.61	3.05
Std. Dev.	4.632	5.644	5.090	5.338	4.146	3.654
Range	17	23	19	19	13	14
Minimum	20	19	-4	19	24	-6
Maximum	37	42	15	38	37	8

Table 2 ANOVA for the difference in scores, separated by quintiles, for phase one.

Quintile	**Sum of the squares**	**Root mean square**	**F**	**Sig.**
1	0.4091	0.4091	1.385	0.36
2	0.5	0,5	3	0.333
3	6	3	0	1
4	0.6944	0.6944	25	0.03777
5	56.25	56.25	225	0.00442

Table 3 Descriptive statistics for both groups in phase two.

	Tablet Group		**IPC Group**	
	Pre-test	**Post-test**	**Pre-test**	**Post-test**
N (students)	23		23	
Min	11	15	13	13
Max	24	26	25	26
Average	17.652	20.130	19.391	19.478
St. Dev.	3.432	2.833	3.407	2.936

Table 4 Evolution of student behavior (number of observations per student) for each session, for both groups in phase two.

	Session 1		Session 2		Session 3	
	Tablet	**IPC**	**Tablet**	**IPC**	**Tablet**	**IPC**
Interrupting	0.47	0.11	0.06	0.03	0.40	0.00
Distraction without causing interruptions	2.63	0.66	1.56	0.23	0.12	0.07
Talking about another topic	0.68	0.05	0.53	0.00	0.64	0.00
Talking about the topic of the session	3.79	0.79	3.39	0.54	4.12	1.13
Talking about the software	2.74	0.39	1.75	0.72	0.76	2.13