

Panorama sobre los sistemas de evaluación automática en línea en matemáticas

Jorge Gaona

jgaonap@docentes.academia.cl

<https://orcid.org/0000-0001-6367-529X>

Universidad Academia de Humanismo Cristiano (UAHC, Chile)

Recibido: 29/04/2020 **Aceptado:** 04/07/2020

Resumen

En este artículo se hace una revisión sistemática de la literatura que trata sobre sistemas de evaluación en línea desde una perspectiva instrumental. Se analiza el contexto en el que se implementan estos sistemas: país, nivel educacional, tema matemático y características de los sistemas. Conjuntamente, se categorizan en tres focos: estudiantes, profesores y artefacto. Los resultados muestran que hay una gran variedad de sistemas y que los trabajos se focalizan (i) en el impacto en los estudiantes: rendimiento, variables socio-afectivas y formas de trabajo; y (ii) en los artefactos: atributos del artefacto, características del artefacto cuyo usuario es el profesor y características de las tareas mediadas por estos sistemas.

Palabras clave: Tecnología; Evaluación online; Tarea; Profesor; Estudiante.

Overview of online assessment systems in mathematics

Abstract

In this paper we make a systematic review of the literature dealing with online assessment systems from an instrumental perspective. The context in which these systems are implemented is analyzed: country, educational level, mathematical subject, and characteristics of the systems. Together, they are categorized into three focuses: students, teachers, and artefacts. The results show that there is a great variety of systems and that the work is focused (i) on the impact on students: performance, socio-affective variables and ways of working; and (ii) on the artefacts: attributes of the artefact, characteristics of the artefact whose user is the teacher and characteristics of the tasks mediated by these systems.

Keywords: Technology; Online assessment; Task; Teacher; Student.

Visão geral dos sistemas de avaliação automática on-line em matemática

Resumo

Este artigo faz uma revisão sistemática da literatura que trata dos sistemas de avaliação em linha a partir de uma perspectiva instrumental. O contexto em que estes sistemas são implementados é analisado: país, nível educacional, disciplina matemática, e características dos sistemas. Juntos, são categorizados em três focos: estudantes, professores e artefatos. Os resultados mostram que existe uma grande variedade de sistemas e que o trabalho se concentra (i) no impacto nos estudantes: desempenho, variáveis socio-afectivas e formas de trabalho; e (ii) nos artefatos: atributos do artefato, características do artefato cujo utilizador é o professor e características das tarefas mediadas por estes sistemas.

Palavras-chave: Tecnologia; Avaliação on-line; Tarefa; Professor; Estudante.

Introducción

En la enseñanza y aprendizaje de la matemática la variedad de tipos de tecnología que se utilizan se ha incrementado en el último tiempo, prueba de ello es la cantidad de libros o capítulos de investigación dedicados exclusivamente al tema, por ejemplo, de las serie *Advances in Mathematics Education*, los estudios de la Comisión Internacional de Instrucción Matemática (ICMI, por sus siglas en inglés), o la serie de libros *Mathematics Education in the Digital Era* de la editorial Springer (Aldon et al., 2017; Aldon & Tragalová, 2019; Clark-Wilson et. al., 2014; Hanna et al., 2019; Celia Hoyles & Lagrange, 2010; Monaghan et al., 2016). Frente a la gran diversidad de soluciones tecnológicas y las formas de uso, parece ser que, no tiene mucho sentido preguntarse si la tecnología sirve o no para aprender matemáticas, más bien, hay que cuestionar si una tecnología específica sirve para uno o varios propósitos. Sin embargo, hay que reconocer que la inversión en infraestructura y formación que han realizado muchos países no se ha traducido en mejoras significativas en el aprendizaje de las matemáticas tal como señala la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2015). Lo anterior nos obliga a sopesar los costos (materiales, instrumentales y temporales) y el valor pragmático y epistémico (Artigue, 2002) que nos suministra el uso de una tecnología específica, junto con otros factores que pueden dificultar o facilitar la integración de estas soluciones en el quehacer cotidiano de los profesores (Gaona, 2018c).

Una de las tecnologías que ha tenido un crecimiento sostenido en el tiempo es aquella que se utiliza para evaluar las matemáticas de forma automática vía internet. En varios libros de educación matemática o sobre tecnología se dedican capítulos específicos para tratar este tópico. Algunos de los cuales dan un panorama general y discuten sobre desafíos técnicos que deben sortear y sus implicancias matemáticas, como el uso de un sistema de álgebra computacional (CAS, por sus siglas en inglés) para la validación de respuestas equivalentes (Sangwin et al., 2010, p. 231). Otros autores, como Stacey y Wiliam (2013, p. 723) discuten sobre las potencialidades pragmáticas -como las posibilidades de tener mucha información a bajo costo- o las posibilidades epistémicas, como la presentación de simulaciones o elementos gráficos específicos. Todo esto guiándose por tres principios. El de las matemáticas: definir cuáles son las matemáticas que los estudiantes deben aprender. El del aprendizaje: crear sistemas que tengan como objetivo mejorar el aprendizaje y el principio de la equidad: apoyar a cada estudiante a aprender las matemáticas que son importantes y demostrar este aprendizaje. En estos principios subyace una idea de evaluación que va más

allá de ser solo sumativa, donde la evaluación se entiende como un proceso de recogida y devolución de información hacia el estudiante, donde el fin está en el aprendizaje (William, 2011).

No obstante, a pesar del entusiasmo por las posibilidades que ofrece esta tecnología, el estudio del uso de sistemas de evaluación en línea en clases ordinarias revela que las tareas tienden a ser muy similares a las tareas en lápiz y papel (Gaona, 2018a, p. 214). Esto no se da solo en ambientes cerrados como las evaluaciones en línea, sino que sucede asimismo con tecnologías de exploración más abiertas (Abboud-Blanchard, 2014, p. 306).

Otras investigaciones más recientes, han ido abordando problemas específicos en ambientes de geometría dinámica, donde algunos sistemas dan una valoración automática a construcciones geométricas o a bosquejos realizados por los estudiantes (Olsher, 2019), o para ayudar de forma automática el proceso de demostración en geometría, con un análisis muy fino de cuáles son los posibles bloqueos de los estudiantes, pero con una cantidad pequeña de tareas automatizadas (Richard et.al., 2019, p. 154).

En la misma línea de lo anterior, hay sistemas que han puesto el acento en la parametrización y que permiten a los profesores crear sus propias preguntas, donde se facilita una retroalimentación en función de la respuesta del estudiante (Kobylanski, 2019) o del proceso de resolución del estudiante (Sangwin, 2019).

Estos trabajos tienen resultados muy interesantes, porque analizan dificultades técnicas y las implicaciones didácticas de estos sistemas. En este artículo se pretende ampliar el análisis de esta tecnología mediante una revisión sistemática de la literatura que trate sobre los sistemas de evaluación automática en línea en matemáticas.

Para nuestros propósitos consideramos un sistema de evaluación en línea en matemáticas como, un artefacto, en el sentido de Rabardel (1995); un artefacto que permite, mediante un dispositivo conectado a internet, recoger las respuestas de los estudiantes, corregirlas automáticamente y guardar parte de la información del trabajo de quien responde.

Bajo esta definición queremos hacer una revisión sistemática de la literatura para explorar, desde una perspectiva instrumental, las siguientes preguntas de investigación: ¿Cuáles son las componentes del artefacto denominado sistema de evaluación en línea? ¿Cuáles son las formas en que se utiliza este artefacto? ¿En qué contextos se utilizan estos artefactos? ¿De qué formas es utilizado el artefacto por los distintos actores: estudiantes, profesores, investigadores y directivos de los centros educacionales?

Marco conceptual

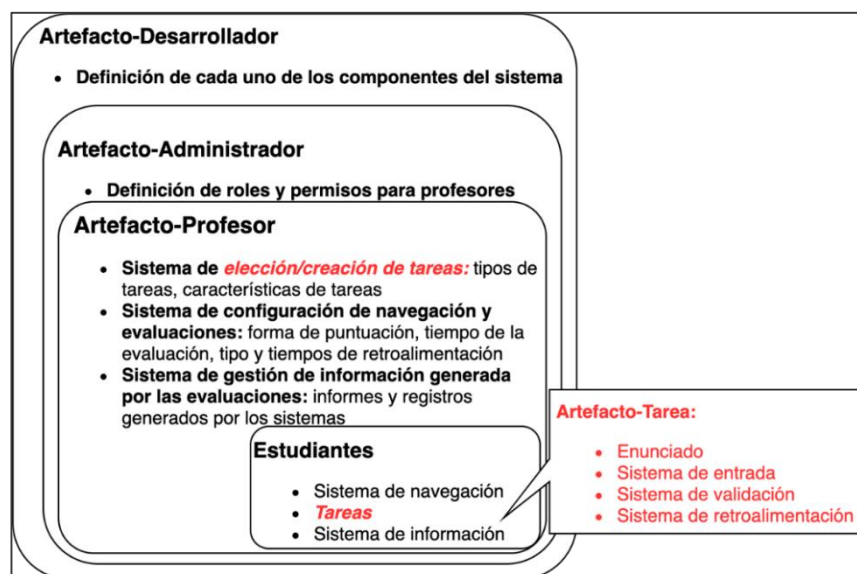
A partir de la perspectiva de (Rabardel, 1995), un artefacto lo designaremos como una construcción material o simbólica con un fin específico. Esta definición excluye al sujeto y la forma en que un sujeto lo utiliza, sin embargo, toma en cuenta las posibilidades de uso que ofrece. De forma general diremos que un artefacto es un conjunto de proposiciones que pueden ser o no usadas por un individuo (Béguin & Rabardel, 2000, p. 17).

Por otra parte, una herramienta es un objeto usado de una forma específica. Por ejemplo, una piedra puede ser usada para cortar o como proyectil y no fue diseñada por el hombre. De este ejemplo se deduce que no toda herramienta es un artefacto.

Finalmente, un instrumento es un artefacto fabricado y usado para algún trabajo técnico específico, como, por ejemplo, un microscopio o un software. Bajo esta definición, el instrumento tiene un componente cognitivo, no es algo dado, sino que es desarrollado por el sujeto, puesto que este debe establecer esquemas de uso (Kuzniak, Nechache, & Drouhard, 2016, p. 864).

A partir de estas definiciones, como cada usuario (desarrollador, administrador, profesor o estudiante) tiene un conjunto de proposiciones distintas sobre las que puede elegir, cada uno puede ser considerado como un artefacto distinto. De forma esquemática se muestran, en la figura 1, diferentes artefactos contenidos uno en el otro, de los cuales cada uno puede ser considerado como un sistema de evaluación en línea particular.

Figura 1: distintos artefactos que podrían ser considerados como un sistema de evaluación en línea y el artefacto-tarea como un componente de cada uno de ellos



Fuente: Elaboración propia (2020)

No es el objetivo de este trabajo definir cada uno de estos artefactos ni sus componentes. Nos centraremos particularmente en las tareas, porque tienen un impacto directo en la actividad matemática potencial de los estudiantes. Los otros componentes, siendo importantes se pueden analizar sin necesariamente tener en cuenta las especificidades de la disciplina, como por ejemplo dar más o menos tiempo para una evaluación específica.

Otro elemento que se desprende de esta definición, es que, dependiendo de las opciones disponibles en cada artefacto, este se puede transformar en un artefacto distinto para dos usuarios en el mismo nivel. Por ejemplo, si en una institución las tareas que componen cada evaluación, la configuración de navegación y evaluaciones vienen definida por un administrador o por un grupo pequeño de profesores, para ese profesor usuario es un artefacto distinto que para aquel que puede definir la cantidad de oportunidades que se le da a los estudiantes o elegir/crear las tareas con las que se evaluarán sus estudiantes.

Las tareas como componentes de un sistema de evaluación en línea, de la misma manera pueden ser consideradas como artefactos en sí mismos porque están diseñadas con un fin específico, el cual es evaluar a un estudiante en matemáticas mediante distintas acciones precisas: visualizar un gráfico, resolver una ecuación, identificar un concepto, arrastrar construir un objeto geométrico que cumpla cierta condición, entre muchas otras posibilidades.

Las características de los componentes del artefacto-tarea, pueden inducir una actividad específica, transformándolas en diferentes instrumentos, por ejemplo, en un instrumento para obtener buenas calificaciones y/o para aprender, para realizar una cierta actividad matemática u otra. Para ilustrar esto podemos remitirnos al estudio realizado por Sangwin y Jones (2017) quienes proponen a los estudiantes dos tipos de tareas en álgebra, directas o reversibles y las aplican en dos formatos, opción múltiple y respuesta abierta. Todo en un sistema de evaluación automática. Los estudiantes obtienen mejores resultados en preguntas de opción múltiple. La diferencia se acentúa cuando la pregunta es reversible. Por ejemplo, cuando al pedirles resolver una ecuación pueden verificar la solución de una lista de opciones versus cuando tienen que escribir la solución en un espacio destinado para ello. En términos de instrumentos, podemos establecer, a partir de los resultados, que el instrumento está midiendo de forma distinta la actividad del estudiante y podemos conjeturar, por lo demás, que la actividad matemática es distinta, por lo que se pone en duda la confiabilidad y validez del instrumento.

Para poder analizar de mejor forma el artefacto-tarea dentro de un sistema de evaluación en línea distinguiremos el conjunto de proposiciones disponibles en cada uno de sus componentes, a saber:

- 1) Enunciado: es donde se propone la acción a realizar por quien responde. En este componente puede haber elementos fijos o parametrizados, estáticos o dinámicos.
- 2) Sistema de entrada: es el componente que permite ingresar la respuesta a quien responde. Podría ser una lista seleccionable (opción múltiple), uno o varios espacios para ingresar números o expresiones algebraicas, uno o varios deslizadores para elegir una posición, un espacio para construir o dibujar un gráfico o un objeto geométrico.
- 3) Sistema de validación: es el componente que compara la respuesta ingresada con la o las respuestas definidas por el estudiante. El sistema podría comparar la respuesta ingresada con la predefinida como correcta de forma literal (ejemplo, $1/2$ es distinto de 0.5 o $x + y$ es distinto de $y + x$); mediante un sistema de cálculo simbólico que distinga una expresión equivalente, una característica específica como si está simplificada o no, o una condición que cumpla un objeto con una característica específica. De igual forma, podría hacer la comparación mediante un procesador geométrico que compare características geométricas como posición, o medidas de distancias, áreas o volúmenes entre otros (ver por ejemplo Gaona & Marquès (2018)).
- 4) Sistema de retroalimentación o feedback: es la información automática que proporciona el sistema a los estudiantes una vez que quien responde ingresa una respuesta. Podría indicarle si su respuesta es correcta o no, darle un feedback en función de la respuesta ingresada (Cazes & Vandebrouck, 2008, p. 187), una recomendación general, una solución paso a paso de la tarea solicitada o una retroalimentación no tan solo de la respuesta final, sino que del proceso de solución del estudiante.

Debemos realizar dos alcances, el primero es que hay veces que dos de estos componentes están unidos en uno solo, por ejemplo, en tareas en ambientes de geometría dinámica el deslizador puede ser parte del enunciado y a su vez el sistema de entrada. El segundo, es que en cada uno de estos componentes podría tener características o configuraciones que son de naturaleza didáctica. Por ejemplo, los tipos de tareas, los objetos matemáticos involucrados y sus representaciones, el enfoque de las tareas (resolución de

problemas, modelación, entre otros), las habilidades que se buscan promover, la validez o confiabilidad del artefacto sistema de evaluación en línea.

Hay que notar que estas configuraciones no son exclusivas de los sistemas de evaluación en línea y sin duda serán tomadas en cuenta, pero en relación a los desafíos, limitaciones y potencialidades que presentan para cada uno de los componentes, del artefacto-tarea, antes mencionados.

Otros artefactos que son de interés para el análisis de la literatura que se realizará en este artículo es el que tiene disponible el profesor y los desarrolladores. Del primero nos interesa particularmente la posibilidad que tienen los profesores de elegir y diseñar tareas (Gaona, 2018b, 2018a). Del segundo, nos interesa cuáles son las elecciones de quienes desarrollan los sistemas y que posibilitan o limitan cierto trabajo matemático.

Método

La revisión de la literatura se realizó mediante una revisión sistemática (Sureda-Negre & Comas-Forgas, 2017). La búsqueda se hizo en Scopus y Web of Science (WOS) para artículos de revistas en idioma español, portugués, francés e inglés publicados entre el 2015 y junio del 2020. En el caso de WOS se seleccionaron en los índices SCI-EXPANDED y SSCI.

- Para el término “evaluación” se buscaron en los títulos de los artículos las siguientes palabras separadas por “or”: “assessment”; “evaluation”; “examination”; “homework”; “quiz”; “test”; “question” y “bank”.
- Para el término “en línea” se buscaron en los títulos de los artículos las siguientes palabras separadas por “or”: “on-line”; “e-exercices”; “e-assessment”; “computer”; “cba”; “caa”; “automatic”; “online”; “digital”; “web” y “technology”.
- Además, se agregaron a la lista de búsqueda palabras relativas a la disciplina en el título, todas separadas por “or”: “arithmetic”; “algebra”; “math*”; “calculus”; “geometry” y “statistics”.
- Finalmente se agregó a la lista de búsqueda palabras relativas al área de educación en el título, palabras clave o resumen, todas separadas por “or”: “stud*”; “teach*”; “learn*”; “education” y “didactic”.

En WOS se encontraron 59 resultados y en Scopus 107 resultados. Se eliminaron los duplicados y quedaron 59 en WOS y 53 de Scopus. De esta lista se buscaron los artículos

que trataran efectivamente sobre sistemas evaluación en línea. Por ejemplo, se excluyó “Active Learning in Undergraduate Mathematics Tutorials Via Cooperative Problem-Based Learning and Peer Assessment with Interactive Online Whiteboards” porque a pesar de contener todas las palabras claves no trata sobre sistemas de evaluación en línea según la definición que adoptamos en este artículo. Al hacer este filtro, quedaron 20 artículos de WOS y 27 de Scopus, totalizando 47 artículos para analizar.

Luego, se analizaron una serie de dimensiones que se separaron en dos partes:

- Contexto: país donde se hizo el estudio, nivel educacional en el que se aplica y tema matemático tratado. En el tema matemático cuando se aplica a una gran variedad de temas se etiquetará como “general”.
- Foco de los artículos: se clasificó según si están centrados en el artefacto en sí mismo (en la multiplicidad de significados que puede tomar, según la figura 1), en los que están centrados en los sistemas de evaluación como un artefacto para los estudiantes y por último para los profesores.

Después de esta clasificación se hizo un análisis particular, estudiando las potencialidades, limitaciones y desafíos que surgen en los componentes de la tarea-artefacto y del artefacto sistemas de evaluación en línea en relación al trabajo matemático potencial de los estudiantes.

Resultados

Los artículos encontrados se clasificaron por nivel educacional y foco, lo que se resume en el cuadro 1.

Cuadro 1: Clasificación de artículos por nivel educacional y por foco de los artículos

Nivel educacional	Estudiantes	Artefacto
Superior	19: (Acosta-Gonzaga & Walet, 2018; Babaali & Gonzalez, 2015; Berková, 2017; Callahan, 2016; Calm et al., 2017; Chow, 2015; Figueroa-Canas & Sancho-Vinuesa, 2017; Gaona, Reguant, Valdivia, Vásquez, & Sancho-Vinuesa, 2018a; Habre & Abboud, 2006; Hauk, Powers, & Segalla, 2015; Jonsdottir, Bjornsdottir, & Stefansson, 2017; Martins, 2017, 2018; Massing, Schwinning, Striewe, Hanck, & Goedicke, 2018; Ocaña, Pulido, Gil, & Zuluaga, 2019; Rønning, 2017; Shotwell & Apigian, 2015; Vasko, Ritter, & Metzger, 2018)	11: (Calvo et al., 2019; Greenhow, 2015; Gwynllyw, Weir, & Henderson, 2016; Herbert, Demskoi, & Cullis, 2019; Maharaj & Dlomo, 2018; Pacheco-Venegas, López, & Andrade-Aréchiga, 2015; Pelkola, Rasila, & Sangwin, 2018; Ramamurthy, Kousika, Mohana, & Mageswari, 2017; Sikurajapathi, Henderson, & Gwynllyw, 2020; Smolinsky, Olafsson, Marx, & Wang, 2019; Smolinsky, Marx, Olafsson, & Ma, 2020)

Secundaria	6: (Bartelet, Ghysels, Groot, Haelermans, & van den Brink, 2016; Fyfe, 2016; Murphy, Roschelle, Feng, & Mason, 2020; Roschelle, Feng, Murphy, & Mason, 2016; Sarmiento, 2017; Walkington, Clinton, & Sparks, 2019)	4: (Gergelitsová & Holan, 2016; Luz & Yerushalmy, 2019; Olsher, Yerushalmy, & Chazan, 2016; Sangwin, 2015)
Primaria	2: (Faber, Luyten, & Visscher, 2017; García, Betts, González-Castro, González-Pienda, & Rodríguez, 2016)	4: (Ginsburg, Lee, & Pappas, 2016; Kiili & Ketamo, 2018; Murchan & Oldham, 2017; Pásztor, Molnár, & Csapó, 2015)
General		1: (Hoogland & Tout, 2018)

Fuente: Elaboración propia (2020).

Un primer resultado sorprendente, a partir de la clasificación que se hizo de los artículos es que no hay trabajos cuyo foco sean los profesores. Esto no quiere decir que los profesores no se nombren en los artículos o no se hable de los potenciales beneficios o dificultades que pueden tener los docentes al usar estos artefactos, pero no hay un análisis sistemático de los profesores como sujetos protagonistas. Esto no quiere decir que no exista literatura al respecto, un ejemplo de esto es el trabajo de Abboud-Blanchard (2014) o Gaona (2018a). En el primero se analiza el uso que hacen los profesores en contextos de clases ordinarias y que está citada en la introducción. En el segundo se hace un análisis de los profesores diseñando tareas en un sistema de evaluación en línea. Sin embargo, en esta búsqueda específica no aparecen y al ampliar la búsqueda, los resultados no son mucho más abundantes.

Contextos

La información sobre los niveles educacionales y el país donde se implementa el sistema de evaluación en línea y el tema matemático que se aborda se resume en el cuadro 2.

Cuadro 2: resumen de elementos contextuales de los sistemas de evaluación en línea. Cada porcentaje está calculado sobre N=47

Nivel educacional/ tema matemático	N	%	Países donde se implementa
Superior	30	63,8%	Alemania (2), Australia, Chile, Colombia, España (3), Estados Unidos (6), Finlandia, India, Inglaterra (5), Islandia, Jamaica, El Líbano, México, Noruega, Portugal (2), República Checa y Sudáfrica
Cálculo	13	27,7%	
General	9	19,1%	
Estadística	6	12,8%	
Álgebra	2	4,3%	
Secundaria	10	21,3%	Estados Unidos (4), Filipinas, Holanda, Inglaterra, Israel (2) y República Checa
General	6	12,8%	

Res. de prob.	1	2,1%	
Inecuaciones	1	2,1%	
Álgebra	1	2,1%	
Geometría	1	2,1%	
Primaria	6	12,8%	España, Estados Unidos, Finlandia, Holanda, Hungría e Irlanda
General	2	4,3%	
Álgebra	1	2,1%	
Res. de prob.	1	2,1%	
Creatividad en mat.	1	2,1%	
Fraciones	1	2,1%	
General	1	2,1%	Sin especificar
Total general	47	100,00%	

Fuente: Elaboración propia (2020).

Podemos observar que la gran mayoría de los artículos (63,8%) se desarrollan en contextos de Educación Superior y dentro de este segmento, una parte importante tratan sobre cálculo diferencial, integral o multivariable, los cuales fueron agrupados bajo la etiqueta cálculo. También, hay una serie de investigaciones que tratan sobre varios temas matemáticos al mismo tiempo. En general, en estos artículos se reporta el uso de sistemas de evaluación en línea que tienen que tener un equilibrio entre las posibilidades que ofrecen y la especificidad que deben tener para evaluar un tema matemático particular. De la misma forma, aparece un número importante de trabajo en estadística y como veremos más adelante, hay varios artefactos diseñados para trabajar con esta disciplina. Por último, hay dos trabajos sobre álgebra, uno de los cuales trabaja con un sistema que evalúa no tan solo de la respuesta final, sino que del proceso de solución del estudiante.

El siguiente nivel donde más se usan estos sistemas es en educación secundaria con un 21,3%. En este segmento aparecen trabajos que discuten las especificidades de diseñar y evaluar tareas sobre temas matemáticos como inecuaciones, geometría o álgebra.

En mucho menor medida se reportan artículos relacionados con el uso de un sistema de evaluación en línea en primaria. Al igual que en el caso de secundaria, se discuten temáticas específicas como, por ejemplo, evaluar automáticamente la creatividad en matemáticas.

La disminución de la cantidad de trabajos a medidas que desciende la edad de los evaluados podría ser por varios motivos. Una suposición que hacemos es que los desafíos

instrumentales que implica trabajar habilidades matemáticas a temprana edad mediante dispositivos de este tipo no son tan sencillos de superar.

Finalmente, hay un artículo que no hace referencia a un sistema educacional específico y sin ser una revisión bibliográfica, proporciona un panorama general de los desafíos que presenta el uso de los sistemas de evaluación en línea en matemáticas.

Otro elemento que destacamos en el cuadro 2, es la gran variedad de países donde se reportan el uso de sistemas de evaluación en línea, específicamente en 27. Esto habla de la penetración de este tipo de tecnología. Particularmente en Latinoamérica se reportan usos en Chile (Gaona, Reguant, Valdivia, Vásquez, & Sancho-Vinuesa, 2018b), Colombia (Ocaña et al., 2019) y México (Pacheco-Venegas et al., 2015). Todos en educación superior. Entendemos que estos artículos son solo una muestra del total de implementaciones que se realizan, puesto que no todas se publican, o no todas están en los índices en que se hizo esta búsqueda y por su puesto debe haber artículos que traten sobre el tema, pero que no se ajustaron a los criterios con los que se hizo este trabajo.

Foco en los estudiantes

En este apartado se especifican los principales resultados donde el foco está en el impacto en los estudiantes del uso de un sistema de evaluación en línea. Se evidencian trabajos que reportan impactos en el rendimiento, en variables socio-afectivas, en las formas de trabajo y en una combinación de estas variables.

Impacto en rendimiento

Hay una serie de trabajos en escuelas secundarias, en las que todos usan el mismo sistema (ASSISTments). Estas investigaciones reportan resultados positivos en las escuelas que usaron el sistema versus aquellos que no lo hicieron. Igualmente mostraron que el efecto fue mayor en estudiantes de bajo rendimiento (Murphy et al., 2020; Roschelle et al., 2016). Utilizando el mismo sistema de evaluación se hicieron variaciones lingüísticas en resolución de problemas y no se obtuvieron diferencias significativas en el desempeño (Walkington et al., 2019). Pero, sí hay efectos positivos para estudiantes de bajo rendimiento cuando se suministra una retroalimentación y el efecto es mayor cuando se provee la respuesta versus cuando se suministra una explicación o cuando se dan más oportunidades. En cambio, en estudiantes de alto rendimiento no hay efecto cuando se provee una retroalimentación (Fyfe, 2016). En estos trabajos, todos usan el mismo artefacto-estudiante, de las distintas

descripciones se puede constatar que el sistema cuenta con aleatoriedad en los enunciados, respuestas numéricas y opción múltiple, no aparece información sobre el sistema de validación y se da una retroalimentación indicando si la respuesta es correcta o incorrecta, y una explicación paso a paso.

En educación superior hay resultados positivos en el uso de estos sistemas en matemáticas generales de primer año (Lowe, 2015; Vasko et al., 2018), en cálculo (Babaali & Gonzalez, 2015; Figueroa-Canas & Sancho-Vinuesa, 2017) y estadística de primer año (Jonsdottir et al., 2017). En estas implementaciones se aprecia que los estudiantes tienen una mayor autonomía para decidir si utilizan o no estos sistemas. Además, se utilizan distintos softwares: iCMA, Stack/Maple T.A, Hawke Learning System y Moodle/Wiris-Quizzes como sistemas de evaluación en línea. En términos de los componentes del artefacto-tarea se describen enunciados con parámetros aleatorios, editores de ecuaciones en los sistemas de entrada (y un sistema de reconocimiento de escritura a mano alzada en el caso de Moodle/Wiris-Quizzes), sistema CAS para la validación y una retroalimentación paso a paso para todos los sistemas salvo Hawke Learning System. De este sistema se reporta el uso de parámetros aleatorios y el uso de respuestas numéricas y retroalimentación que indica correcto/incorrecto, no se mencionan características sobre el sistema de validación.

Siguiendo en educación superior, hay otros trabajos que indican que no hay diferencias significativas entre aquellos que usan un sistema de evaluación en línea versus aquellos que no, tanto en matemáticas generales (Chow, 2015) como en álgebra (Hauk et al., 2015). Para el primero no hay información del artefacto-tarea y para el segundo reportan el uso de un sistema llamado WebWork, el cual tiene parámetros aleatorios en los enunciados, respuestas abiertas para números y expresiones algebraicas sin editor, CAS como sistema de validación y retroalimentación que indica correcto/incorrecto.

Impacto en variables socio-afectivas

En secundaria el uso de MyOpenMath los motiva y les proporciona confianza en las matemáticas. Este sistema tiene parámetros aleatorios en los enunciados, respuestas numéricas y algebraicas, sin editor para el ingreso, pero con un sistema que da una vista previa antes de enviar la respuesta para verificar si está bien escrito. También, cuenta con un CAS para evaluar expresiones equivalentes. Por último, tiene un sistema de ingreso geométrico para dibujar algunos objetos como rectas o puntos.

En educación superior hay dos trabajos, uno en cálculo y otro en matemáticas generales. En ambos hay impactos positivos en variables socio-afectivas: en la utilidad percibida utilizando Maple T.A (Rønning, 2017) y utilidad y agrado percibidos utilizando STACK (Acosta-Gonzaga & Walet, 2018). Ambos sistemas tienen varios elementos en común: enunciado parametrizable, respuestas numéricas y algebraicas, CAS para la evaluación. Las diferencias son que Maple T.A tiene un editor de ecuaciones para el ingreso de respuestas y STACK un sistema de vista previa al igual que MyOpenMath y que el primero solo indica si la respuesta ingresada es correcta/incorrecta y en el segundo se puede diseñar una retroalimentación paso a paso. Esta diferencia produce críticas de los estudiantes que solo reciben como feedback correcto o incorrecto, en cambio, recibir una retroalimentación sofisticada es una de las características más apreciadas por los estudiantes.

Impacto en el trabajo de los estudiantes

En esta categoría se clasificaron dos artículos. El primero se desarrolla en educación primaria y se reportan el análisis automático de distintas estrategias revisadas de forma automática. En este trabajo el acento técnico está en la capacidad del software en la detección de estrategias de resolución en dos tareas de resolución de problemas (García et al., 2016). El otro trabajo reportado es en educación superior en un curso de nivelación y se usa Moodle/Wiris-Quizzes (Gaona et al., 2018a). Se analiza la diferencia en el trabajo de los estudiantes al variar el momento en que se suministra la retroalimentación paso a paso: inmediatamente o diferida (una vez terminado el cuestionario). Se concluyó que esta última ayuda a optimizar el trabajo de los estudiantes puesto que distribuye el trabajo en todo el cuestionario y reduce los intentos sin que disminuya la calificación.

Impacto en una combinación de las variables anteriores

En esta categoría se encuentran estudios que analizan en forma combinada dos de las variables antes mencionadas: rendimiento, variables socio-afectivas y formas de trabajo.

Los artículos que analizan rendimiento y variables socio-afectivas son, en su mayoría, en educación superior. Algunos de estos trabajos reportan un impacto positivo en el rendimiento y en la satisfacción de estudiantes (Calm et al., 2017; Martins, 2017, 2018), todos en cálculo y cálculo multivariable. Mientras que Callahan (2016) reporta impacto positivo en la retención y participación, pero no en el rendimiento. Los trabajos mencionados acá utilizan sistemas ya especificados más arriba, como Moodle/Wiris-Quizzes y WebWork,

salvo Martins (2017;2018) que trabaja solo con Moodle y Latex lo que lo obliga a fijar los enunciados y trabajar solo con preguntas de opción múltiple.

Solo el trabajo de Faber et al (2017) se inscribe en educación primaria, donde reporta un aumento en la motivación y el rendimiento, con un impacto mayor para aquellos que son considerados como buenos alumnos. Resultado totalmente distinto a los encontrados en educación secundaria. Lo curioso es que usan ASSISTments, así que lo único particularmente distinto es el nivel donde se aplica el sistema.

Finalmente hay otro grupo de artículos que analizan el impacto en el rendimiento y en las formas de trabajo de los estudiantes, todos en educación superior. Están los que reportan impactos positivos en el rendimiento y que, además, exploran las razones por la cual los estudiantes realizan una nula, media o alta cantidad de intentos en un curso de matemática general (Ocaña et al., 2019), en la nula correlación entre enfoques de aprendizaje y el rendimiento, aunque se constata que aquellos estudiantes con un enfoque de aprendizaje profundo necesitan menos intentos en un curso de cálculo (Berková, 2017) o una correlación positiva entre tiempo de trabajo en estadística (Massing et al., 2018). En el primero se utiliza solo Moodle lo que limita las posibilidades de la tarea-artefacto a preguntas de opción múltiple, en el segundo utiliza Maple T.A y en el tercero una combinación de un LMS con R, que permite parametrizar los enunciados, dar una retroalimentación general y trabajar con comandos específicos de estadística.

Por otra parte, están los artículos que no reportan impacto positivo en el rendimiento, pero muestran como estos sistemas pueden ser utilizados para detectar a estudiantes particularmente débiles en cálculo (Habre, 2015) o que son los deberes en línea los que inicialmente comienzan a trabajar los estudiantes entre un conjunto de otros recursos digitales como libros o videos (Shotwell & Apigian, 2015). El primero usa MyMathLab que permite parametrizar enunciados, utilizar un editor de ecuaciones para el ingreso de expresiones matemáticas e indicar si la respuesta correcta/incorrecta, pero no aparece información sobre el uso de un CAS para evaluar equivalencias u otras propiedades.

Foco en el artefacto

En este apartado aparecen resultados que se centran principalmente en el artefacto, particularmente sus potencialidades y limitaciones en relación a artefacto-estudiante, artefacto-profesor, artefacto-tarea o (Ver figura 1).

Atributos del artefacto-profesor

Hay una serie de artículos que trabajan sobre atributos del artefacto-profesor y no sobre un componente particular, específicamente sobre su validez y/o confiabilidad.

En primaria se estudia la confiabilidad de una evaluación a través del juego para fracciones y, además, se mostró que redujo la ansiedad de niños y niñas que lo utilizaron (Kiili & Ketamo, 2018). Asimismo, se probó la validez y confiabilidad de un sistema de evaluación que, igualmente, permite automatizar la recolección de información sobre el pensamiento matemático para los profesores, por ejemplo, las estrategias que utilizan los niños y niñas para resolver un problema de aritmética (Ginsburg et al., 2016).

En educación superior Maharaj y Dlomo (2018) comprueban que una evaluación diagnóstica en línea es un buen predictor de los resultados finales, sin embargo no se indica el software que se utiliza ni sus características. Smolinsky et al. (2019; 2020) constatan que un sistema de evaluación en línea es válido y confiable teniendo como parámetro de comparación las evaluaciones en papel en un curso presencial de cálculo y en un curso online de cálculo utilizando el sistema WebAssign. Resultados distintos a los obtenidos por Pelkola et al. (2018) quienes concluyen que las evaluaciones en línea en cálculo, usando STACK, todavía no son un sustituto de las evaluaciones en papel. Parece ser que las diferencias están en el enfoque metodológico entre los dos últimos estudios, puesto que, en términos de artefacto, STACK parece ser más sofisticado que WebAssign.

Artefacto-profesor

En esta categoría clasificamos el artículo de Calvo et al. (2019), quien trabaja en un sistema para generar trazas de la cantidad de trabajo de los estudiantes, los caminos tomados y la corrección de la respuesta específica. Aunque el último de estos indicadores pertenece al artefacto-tarea, acá se busca no tan solo indicar si una respuesta dada por los estudiantes es correcta o no, sino que dar información global de un conjunto de respuestas. De la misma forma, establecen la confiabilidad de estas correcciones en comparación con evaluaciones.

Artefacto-tarea

El primer artículo que presentamos en esta categoría es el de Hoogland y Tout (2018), quienes hacen un análisis general de los sistemas tomando en cuenta las presiones políticas,

sociales, educacionales y tecnológicas. Luego, ofrece un marco para analizar las tareas de evaluación matemática, donde establece una jerarquía creciente de sofisticación y luego distintas categorías como: tipos de representación (descontextualizados, problemas contextualizados simples, contextualización más compleja, sofisticados problemas contextuales estáticos de representación e interactividad en el espacio de respuesta, pero no interactivos en el espacio de situación, entre otros), posible entrega (lápiz/papel o sistema de evaluación en línea), posibles tipos de respuesta (opción múltiple, respuesta abierta numérica, respuesta abierta algebraica o sistema arrastrar/soltar) y habilidades del siglo 21 (desde poco a mucho). A pesar de que los elementos que muestra son interesantes, en las categorías se mezclan elementos que son de naturaleza completamente distinta. Por ejemplo, en los tipos de representación habla de interactividad en el ingreso y también, de problemas no contextualizados. La matemática como construcción humana puede generar tareas que requieran una serie de habilidades complejas, por lo que la representación por si sola no podría clasificar la habilidad que promueve. En este sentido, analizar las posibilidades que ofrecen los componentes del artefacto-tarea parece ser más pertinente para analizar un sistema de evaluación en línea.

De acuerdo a la figura 1, el artefacto-tarea tiene distintos componentes: enunciado, sistemas de entrada, sistema de validación y sistema de feedback. Como hay algunos artículos que discuten sobre más de uno y los resultados se presentarán usando estas sub-categorías, por lo que algunos trabajos se pueden repetir.

En Gergelitsová y Holan (2016) trabajan con un sistema de evaluación en geometría que mediante un applet de GeoGebra propone a estudiantes de secundaria tareas con parámetros fijos, que pueden tener más de una solución y donde el sistema registra la respuesta mediante los pasos de construcción. Para validar considera los objetos geométricos definidos como correctos y los compara con los suministrados por los estudiantes. Después provee la corrección de esta evaluación como correcta/incorrecta. Finalmente, dan cuenta de un panel de gestión de las calificaciones del artefacto-profesor. Lo interesante del artículo es que analiza tareas específicas que ejemplifican cómo funciona el sistema, las posibilidades en términos de construcción y justificación. Los autores expresan que las tareas vienen diseñadas y que los profesores no tienen un espacio de creación, esto lo ven como una ventaja puesto que indican falsos negativos en la validación de las respuestas.

Los desarrollos de la plataforma STEP igualmente son en geometría (Luz & Yerushalmy, 2019; Olsher et al., 2016), aunque igualmente los aplican en tareas de cálculo.

Al igual que el artículo anterior, se trabaja con una modificación de GeoGebra y se implementa en escuelas secundarias. En este caso se hace hincapié en tareas donde se entrega una condición que un objeto matemático debe cumplir. En el primer artículo dan como ejemplo una tarea que pide a los estudiantes que interactúen cambiando la posición de los puntos de una construcción para que la proporción de las áreas de dos triángulos esté en la razón 1:4.

En el segundo artículo, a partir de un diagrama compuesto por varios triángulos, se pidió a los estudiantes que utilizaran el arrastre para generar un contra-ejemplo a una declaración universal, un ejemplo demostrativo a una declaración existencial o que declararan que los ejemplos solicitados no existían. En ambos casos se da una clasificación de cada respuesta como correcta/incorrecta pese a la infinidad de posibles respuestas correctas existentes. Además, se hace una clasificación de una serie de errores u objetos correctos con propiedades particulares, todas estas posibles respuestas previstas gracias a investigaciones en didáctica de las matemáticas y al análisis de las propiedades de los objetos matemáticos involucrados. En estos artículos, la diferencia con el trabajo de Gergelitsová y Holan está precisamente en los tipos de tareas (con infinitas soluciones), en un sistema capaz de detectar todas las posibles respuestas automáticamente y en la clasificación de las respuestas en categorías que dan información sobre las concepciones matemáticas de los estudiantes. Esta información se almacena y se entrega mediante un sistema de gestión que está en el artefacto-profesor.

En la misma línea de análisis de errores, pero en educación primaria se constatan una serie de distinciones que hacen en los posibles errores para los estudiantes, en varios problemas matemáticos, como calcular $267c + 131c$ y escribirlos en euros. Estos errores eran previstos a partir de investigaciones previas que categorizaron los errores en evaluaciones hechas en lápiz/papel. Asimismo, agregan errores de ingreso a la respuesta, por ejemplo, presionar involuntariamente la barra espaciadora, de tal forma de dar puntaje completo en estos casos. Una de las discusiones que deja abierta es si, por ejemplo, una respuesta que sea 3.90 es una aproximación adecuada y cuánto puntaje se debería dar por eso, sin embargo, está la posibilidad técnica de categorizarlo, puntuarlo y dar un feedback al respecto.

En educación superior está el trabajo de Sikurajapathi et al. (2020) en cálculo y Gwynllyw et al. (2016) en estadística, quienes hicieron un proceso similar, es decir, se categorizan los errores y se programó una retroalimentación para los errores detectados,

puesto que en este caso se indica explícitamente que los parámetros del enunciado son aleatorios. La diferencia entre ambos trabajos es que cada uno utiliza un software específico para evaluar elementos propios del cálculo con un CAS y la estadística con el software R.

En Pacheco-Venegas et al. (2015) se muestra un sistema con capacidad para ir evaluando el paso a paso de un proceso de solución algebraico, como expandir y simplificar una expresión algebraica o resolver una ecuación. En este proceso, el estudiante a medida que va ingresando líneas de respuestas, el sistema le va indicando si es correcta o no. Las preguntas son fijas, y se desarrolla con un sistema denominado MathDIP.

Todos los trabajos anteriores dependen, para la detección de errores, del sistema de validación. En el trabajo de Sangwin (2015) se discute la capacidad de distintos CAS para calcular inecuaciones, capacidad que luego es usada para generar el sistema de evaluación automática. En este artículo se puede ver que las potencialidades y limitaciones que tenga el sistema de validación incidirá en la forma que tenga luego el sistema de comparar la respuesta correcta con la ingresada y a partir de esto puntuar y suministrar un feedback, similar a lo discutido en Ramamurthy (2017).

En una línea un tanto distinta está el trabajo de Herbert et al. (2019) quienes desarrollan un sistema que genera archivos PDF que pueden ser completados por los estudiantes fuera de una plataforma, son subidos a una plataforma y este los corrige automáticamente. Esto podría ser particularmente interesante para estudiantes que tienen un acceso limitado a una conexión a internet.

Finalmente hay un artículo que habla de una serie de argumentos a favor y en contra para su uso y las capacidades que debería tener el sistema de evaluación (Greenhow, 2015). Por una parte, establece una serie de argumentos a favor para usar estos sistemas que son de naturaleza pragmática: posibilidad de flexibilidad de la evaluación para los estudiantes o disminuir tiempos de corrección de los profesores y otro que podría ser de naturaleza más epistémico, como es la interactividad de los enunciados. En contraparte, muestra una serie de limitaciones, como son la dificultad de dar créditos parciales y la injusticia que esto representa, la imposibilidad de trazar un gráfico como respuesta o la imposibilidad de hacer preguntas matemáticas más complejas donde el estudiante deba realizar y justificar distintos procesos. Finalmente establece una serie de características deseables que deberían tener estos sistemas, entre los cuales se destacan: la posibilidad de parametrizar enunciados, aunque advierte de ciertas consideraciones didácticas de la parametrización, por ejemplo, si en una ecuación cuadrática se dejan de forma aleatoria cada coeficiente, la dificultad

cambiará cuando el determinante sea un número racional o irracional. La posibilidad de detectar errores y retroalimentar en función de ellos y puntuaciones que vayan más allá de 1 y 0.

Conclusiones

En este artículo se hizo una revisión sistemática de la literatura que trata sobre los sistemas de evaluación en línea desde una perspectiva instrumental. Esta aproximación teórica permitió hacer una diferencia entre los distintos artefactos denominados como sistema de evaluación en línea. Una vez realizada esta distinción, se clasificaron los artículos, entre aquellos que tenían como foco 1) los estudiantes, 2) los profesores y 3) el artefacto sistema de evaluación en línea, en sus distintas versiones o el artefacto-tarea. Además, se analizaron elementos contextuales como el país o el nivel educativo donde se llevaron a cabo las experimentaciones. Una de las limitaciones de este trabajo es que, sin duda, no están todos los trabajos producidos sobre el tema en los índices buscados, sin embargo, consideramos que esta exploración es representativa y además puede ser ampliada por otras u otros investigadores.

El análisis reveló la no presencia de artículos donde se analiza al profesor. En cambio, pudimos constatar que hay un equilibrio entre artículos cuyo foco es el artefacto o cuyo foco son los estudiantes, con una predominancia de este último. Al mismo tiempo, hay una disminución de los trabajos a medida que disminuye la edad de los estudiantes, concentrándose la gran mayoría en educación superior.

Cuando el foco está puesto en el estudiante, se reportan resultados mixtos sobre el impacto en el rendimiento, aunque con una predominancia de resultados positivos cuando se utilizan. Se observa que se hacen una serie de cruces de variables como el tipo de retroalimentación que se facilita o el perfil de los estudiantes.

Cuando se estudia el impacto en variables socio-afectivas todos los resultados reportados son positivos. Se estudian efectos sobre la adherencia, retención, ansiedad y satisfacción. En esta última variable se advierte un efecto positivo mayor cuando hay una retroalimentación paso a paso versus una donde solo se indica si la respuesta ingresada es correcta o incorrecta.

Otra categoría que aparece como centro de análisis es el impacto del uso de estos sistemas en el trabajo de los estudiantes, particularmente la cantidad de intentos que realizan,

el tiempo de trabajo y las fluctuaciones de estas variables cuando se modifica el momento en que se entrega esta retroalimentación.

En los estudios en los que el foco es el estudiante se repara que, en general, se utilizan artefactos-tarea que cuentan en los enunciados con elementos aleatorios. No se notó la presencia de simulaciones o elementos dinámicos. Frente a lo anterior, sería particularmente interesante explorar, mediante nuevas investigaciones, si hay diferencias en alguna de las categorías precedentes cuando se utilizan modificaciones entre lo estático y dinámico. Esto se aborda en los artículos cuyo foco son los artefactos. Por ejemplo, en aquellos en aquellos que tratan sobre geometría y donde los enunciados son objetos deslizables.

Lo que no aparece en esta búsqueda, y que puede ser interesante de explorar, es analizar las diferencias para el estudiante respecto a la variabilidad de los parámetros cuando estos son aleatorios. Por ejemplo, cuando hay un problema donde parte de la solución se obtiene al resolver una ecuación del tipo $ax^2 + bx + c = 0$, con a , b y c aleatorios. Si no se controlan los parámetros, para ciertos valores puede no existir solución en los reales, puede tener dos soluciones iguales o dos diferentes. Más aún, para ciertos valores las soluciones pueden ser números enteros, racionales (no enteros) o irracionales ¿qué tan distintos son los efectos en el trabajo de los estudiantes o en el rendimiento frente a estas distintas variaciones?

De la misma forma, a lo anterior se podría agregar como programa de investigación la caracterización del trabajo de los estudiantes a partir de cambios de registros semióticos (Duval, 1995), tipos de tareas, tipos de contextualización, características de los objetos matemáticos involucrados, entre otras posibilidades. Para ejemplificar, si se hace una pregunta sobre un gráfico ¿da lo mismo que los coeficientes de la función sean enteros? Es decir ¿el trabajo, rendimiento o percepción de los estudiantes cambia si las tareas que enfrentan tienen estas variaciones?

En los artículos cuyo foco es el artefacto se demuestra que para tratar ciertos temas particulares de cálculo, álgebra, geometría o estadísticas el trabajo se enriquece cuando se utilizan softwares específicos para esos temas y hay una preocupación por la validez epistemológica de los objetos involucrados. No obstante, esto plantea problemas técnicos para las instituciones que los adoptan: si una escuela o universidad debe impartir cursos en todos esos dominios matemáticos, ¿cuál software utiliza? ¿levanta una plataforma para cada

curso o sacrifica la riqueza del trabajo o la validez epistemológica por una facilidad de ingreso y navegación del estudiante?

En los sistemas de entrada aparecen preguntas abiertas numéricas o algebraicas y hay algunos sistemas que cuentan con editor de ecuaciones, o un sistema de reconocimiento de escritura a mano alzada. Sin embargo, la gran mayoría utiliza texto plano para la escritura de las respuestas o un sistema de vista previa para visualizar lo ingresado por los estudiantes antes de enviar la respuesta. Se aprecia en menor medida el uso exclusivo de preguntas de opción múltiple, aunque siguen siendo utilizadas, pero de forma complementaria. No obstante, son relativamente escasos los trabajos que analizan la validez o la confiabilidad de las pruebas de opción múltiple en matemáticas, aunque existen algunos trabajos fuera de esta búsqueda que tratan el tema, por ejemplo, en el caso de los gráficos (Berg & Boote, 2015).

En los sistemas de validación es interesante observar la evolución de los softwares utilizados para esto. En el caso de los CAS hace unos años el desafío era aceptar como respuesta correcta a dos expresiones equivalentes. En los artículos que se analizaron, se observan sistemas muchos más complejos que permiten ver, más allá de la equivalencia, si una expresión tiene un objeto cumple con una propiedad específica: analizar si la expresión está simplificada, factorizada o si una función cumple tal o cual propiedad. Esta misma potencia se está utilizando para identificar errores típicos o concepciones de los estudiantes. En este sentido, se ha visto un enriquecimiento de los análisis más allá de la dicotomía correcto/incorrecto.

La capacidad de los CAS para validar propiedades y errores debería servir para generar sistemas de retroalimentación más personalizados. Pero, en los artículos explorados, solo se encontró un artículo que estudia el impacto en el rendimiento de los estudiantes cuando se varía la retroalimentación. Estos resultados son coherentes con otros en la misma línea (Attali, 2015; Attali & van der Kleij, 2017). Esto deja una serie de preguntas abiertas: ¿cuáles son las retroalimentaciones más adecuadas para los estudiantes? ¿será positivo agregar distintos registros en una retroalimentación? ¿podrá una retroalimentación específica modificar la forma en la que un estudiante aborda un problema?

Finalmente nos preguntamos, ¿cuál es el papel del profesor en este proceso? En la exploración que realizamos su rol está invisibilizado, pareciera ser solo un usuario, a pesar que el artefacto utilizado tiene una serie opciones sobre las que debe decidir, en particular cuando son los profesores quienes pueden ser diseñadores de tareas. Aunque no aparece en esta revisión, existe una incipiente exploración sobre decisiones de los profesores cuando

estos diseñan tareas en un sistema de evaluación en línea (Gaona, 2018a), considero que esta es también una línea interesante para seguir comprendiendo el alcance de estos sistemas y su impacto en el ecosistema educativo.

Este trabajo, de alguna forma, da cuenta de la heterogeneidad de lo que se denomina un sistema de evaluación en línea, si nos hacemos la pregunta ¿funcionan los sistemas de evaluación en línea en matemáticas? Se debería especificar funciona ¿para qué? ¿Para tener mejor rendimiento? ¿Para mejorar variables socio-afectivas? ¿Para dar retroalimentación? ¿para que los profesores tengan más y mejor información para actuar? Asimismo, se deberían especificar cuáles son las características del artefacto específico que se está trabajando, esto podría ayudar a comprender cuáles son las potencialidades y limitaciones que se están fomentando.

En tiempos donde la virtualidad ha cobrado especial interés es importante cuestionar la tecnología, delimitando lo más posible su campo de acción, sus potencialidades y limitaciones. La investigación tiene que jugar un rol fundamental para responder las preguntas aquí planteadas y otras que vayan en la dirección de cuestionar cuál es el aporte en el proceso de enseñanza y aprendizaje de los estudiantes.

Referencias

- Abboud-Blanchard, M. (2014). Teachers and Technologies : Shared Constraints, Common Responses. En A. Clark-Wilson, O. Robutti, & N. Sinclair (Eds.), *Technology in Mathematics Teaching* (pp. 297–317). Dordrecht: Springer Netherlands. https://doi.org/https://doi.org/10.1007/978-94-007-4638-1_13
- Acosta-Gonzaga, E. & Walet, N. (2018). The role of attitudinal factors in mathematical on-line assessments: a study of undergraduate STEM students. *Assessment and Evaluation in Higher Education*, 43(5), 710–726. <https://doi.org/10.1080/02602938.2017.1401976>
- Aldon, G., Hitt, F., Bazzini, L. & Gellert, U. (2017). *Mathematics and technology*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-51380-5>
- Aldon, G. & Tragalová, J. (2019). *Technology in Mathematics Teaching*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-19741-4>
- Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, 7(3), 245–274. <https://doi.org/https://doi.org/10.1023/A:1022103903080>
- Attali, Y. (2015). Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems. *Computers and Education*, 86, 260–267. <https://doi.org/10.1016/j.compedu.2015.08.011>

- Attali, Y. & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers and Education*, 110, 154–169. <https://doi.org/10.1016/j.compedu.2017.03.012>
- Babaali, P. & Gonzalez, L. (2015). A quantitative analysis of the relationship between an online homework system and student achievement in pre-calculus. *International Journal of Mathematical Education in Science and Technology*, 46(5), 687–699. <https://doi.org/10.1080/0020739X.2014.997318>
- Bartelet, D., Ghysels, J., Groot, W., Haelermans, C. & van den Brink, H. M. (2016). The differential effect of basic mathematics skills homework via a web-based intelligent tutoring system across achievement subgroups and mathematics domains: A randomized field experiment. *Journal of Educational Psychology*, 108(1), 1–20. <https://doi.org/10.1037/edu0000051>
- Béguin, P. & Rabardel, P. (2000). Concevoir pour les activités instrumentées. *Revue d'Intelligence Artificielle*, 14(1–2), 35–54.
- Berg, C. & Boote, S. (2015). Format Effects of Empirically Derived Multiple-Choice Versus Free-Response Instruments When Assessing Graphing Abilities. *International Journal of Science and Mathematics Education*, 15(1), 19–38. <https://doi.org/https://doi.org/10.1007/s10763-015-9678-6>
- Berková, A. J. (2017). Effect of the use of computer-aided assessment system in the teaching of mathematical analysis with regard to students' approaches to learning. *Journal on Efficiency and Responsibility in Education and Science*, 10(3), 71–75. <https://doi.org/10.7160/eriesj.2017.100302>
- Callahan, J. T. (2016). Assessing Online Homework in First-Semester Calculus. *Primus*, 26(6), 545–556. <https://doi.org/10.1080/10511970.2015.1128501>
- Calm, R., Masià, R., Olivé, C., Parés, N., F, P., Ripoll, J. & Sancho-Vinuesa, T. (2017). Use of WIRIS quizzes in an online calculus course. *Journal of Technology and science Education*, 5(3), 184–193.
- Calvo, M., Carnicer, A., Cuadros, J., Martori, F., Miñarro, A. & Serrano, V. (2019). Computer-assisted assessment in open-ended activities through the analysis of traces: A proof of concept in statistics with R commander. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(9). <https://doi.org/10.29333/ejmste/108456>
- Cazes, C. & Vandebrouck, F. (2008). Panorama sur les bases d'exercices en ligne. En F. Vandebrouck (Ed.), *La classe de mathématiques : activités des élèves et pratiques des enseignants* (pp. 183–196). Toulouse: Octarès.
- Chow, A. F. (2015). Online homework impact in undergraduate mathematics and business statistics courses. *Educational Studies*, 41(3), 244–248. <https://doi.org/10.1080/03055698.2014.961902>
- Clark-Wilson, A., Robutti, O. & Sinclair, N. (2014). *The Mathematics Teacher in the Digital Era*. Dordrecht: Springer. <https://doi.org/10.1007/978-94-007-4638-1>
- Duval, R. (1995). *Sémiosis et pensée humaine: registres sémiotiques et apprentissages intellectuels*. Suisse: Peter Lang.
- Faber, J. M., Luyten, H. & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education*, 106, 83–96. <https://doi.org/10.1016/j.compedu.2016.12.001>

- Figuroa-Canas, J. & Sancho-Vinuesa, T. (2017). Exploring the Efficacy of Practicing with Wiris-Quizzes in Online Engineering Mathematics. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 12(3), 141–146. <https://doi.org/10.1109/RITA.2017.2735499>
- Fyfe, E. R. (2016). Providing feedback on computer-based algebra homework in middle-school classrooms. *Computers in Human Behavior*, 63, 568–574. <https://doi.org/10.1016/j.chb.2016.05.082>
- Gaona, J. (2018a). *Elaboración de un sistema de evaluación en línea como proceso de formación de profesores de matemáticas*. Université Sorbonne Paris Cité - Université Paris Diderot. Recuperado de <https://tel.archives-ouvertes.fr/tel-02458946/>
- Gaona, J. (2018b). Instructors' decision making when designing resources: the case of online assessments. En V. Giritana, T. Miyakawa, M. Rafalska, & S. Soury-Lavergne (Eds.), *Re(s)ources 2018 International Conference* (pp. 279–282). Lyon: École Normale Supérieure de Lyon.
- Gaona, J. (2018c). Integrar tecnología en la enseñanza-aprendizaje de las matemáticas, factores claves. *Revista de Gestión de la Innovación en Educación Superior*, 3, 75–93. Recuperado de <http://ojs.inacap.cl/index.php/regies/article/view/102>
- Gaona, J. & Marquès, D. (2018). Développer dans Moodle une base d'exercices mathématiques en ligne. Problématiques croisées entre didacticiens et informaticiens : le cas de Wiris au Chili. *Cahiers du laboratoire de didactique André Revuz*, 19, 23–31. Recuperado de <https://hal.archives-ouvertes.fr/hal-02111650>
- Gaona, J., Reguant, M., Valdivia, I., Vásquez, M. & Sancho-Vinuesa, T. (2018a). Feedback by automatic assessment systems used in mathematics homework in the engineering field. *Computer Applications in Engineering Education*, 26(4), 994–1007. <https://doi.org/10.1002/cae.21950>
- Gaona, J., Reguant, M., Valdivia, I., Vásquez, M. & Sancho-Vinuesa, T. (2018b). Feedback by automatic assessment systems used in mathematics homework in the engineering field. *Computer Applications in Engineering Education*, 26(4), 994–1007. <https://doi.org/10.1002/cae.21950>
- García, T., Betts, L., González-Castro, P., González-Pienda, J. A. & Rodríguez, C. (2016). Evaluación on-line del proceso de resolución de problemas matemáticos en estudiantes de quinto y sexto curso: auto-regulación y logro. *Revista Latinoamericana de Investigación en Matemática Educativa*, 19(2), 165–186. <https://doi.org/10.12802/relime.13.1922>
- Gergelitsová, Š., & Holan, T. (2016). GeoTest - A system for the automatic evaluation of geometry-based problems. *Computer Applications in Engineering Education*, 24(2), 297–304. <https://doi.org/10.1002/cae.21712>
- Ginsburg, H., Lee, Y. S., & Pappas, S. (2016). A research-inspired and computer-guided clinical interview for mathematics assessment: introduction, reliability and validity. *ZDM - Mathematics Education*, 48(7), 1003–1018. <https://doi.org/10.1007/s11858-016-0794-8>
- Greenhow, M. (2015). Effective computer-aided assessment of mathematics; principles, practice and results. *Teaching Mathematics and its Applications*, 34(3), 117–137. <https://doi.org/10.1093/teamat/hrv012>
- Gwynllyw, R., Weir, I. & Henderson, K. (2016). Using DEWIS and R for multi-staged

- statistics e-Assessments. *Teaching Mathematics and its Applications*, 35(1), 14–26. <https://doi.org/10.1093/teamat/hrv018>
- Habre, S. (2015). Effects of web-based homework on students' performance in freshman calculus at an American college in Lebanon. *Turkish Online Journal of Educational Technology*, (Special Issue 1), 220–226.
- Habre, S., & Abboud, M. (2006). Students' conceptual understanding of a function and its derivative in an experimental calculus course. *The Journal of Mathematical Behavior*, 25(1), 57–72.
- Hanna, G., Reid, D. & Villiers, M. de. (2019). *Proof Technology in Mathematics Research and Teaching* (Vol. 14). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-28483-1>
- Hauk, S., Powers, R. A. & Segalla, A. (2015). A Comparison of Web-based and Paper-and-Pencil Homework on Student Performance in College Algebra. *Primus*, 25(1), 61–79. <https://doi.org/10.1080/10511970.2014.906006>
- Herbert, K., Demskoi, D. & Cullis, K. (2019). Creating mathematics formative assessments using LaTeX, PDF forms and computer algebra. *Australasian Journal of Educational Technology*, 35(5), 153–167. <https://doi.org/10.14742/ajet.4539>
- Hoogland, K., & Tout, D. (2018). Computer-based assessment of mathematics into the twenty-first century: pressures and tensions. *ZDM - Mathematics Education*, 50(4), 675–686.
- Hoyles, C. & Lagrange, J.-B. (2010). *Mathematics Education and Technology-Rethinking the Terrain*. New York: Springer US. <https://doi.org/10.1007/978-1-4419-0146-0>
- Jonsdottir, A. H., Bjornsdottir, A. & Stefansson, G. (2017). Difference in learning among students doing pen-and-paper homework compared to web-based homework in an introductory statistics course. *Journal of Statistics Education*, 25(1), 12–20. <https://doi.org/10.1080/10691898.2017.1291289>
- Kiili, K. & Ketamo, H. (2018). Evaluating Cognitive and Affective Outcomes of a Digital Game-Based Math Test. *IEEE Transactions on Learning Technologies*, 11(2), 255–263. <https://doi.org/10.1109/TLT.2017.2687458>
- Kobylanski, M. (2019). WIMS: Innovative Pedagogy with 21 Year Old Interactive Exercise Software. En G. Hanna, D. Reid, & M. de Villiers (Eds.), *Proof Technology in Mathematics Research and Teaching* (pp. 123–144). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-19741-4_6
- Kuzniak, A., Nechache, A. & Drouhard, J. P. (2016). Understanding the development of mathematical work in the context of the classroom. *ZDM - Mathematics Education*, 48(6), 861–874. Recuperado de <http://link.springer.com/10.1007/s11858-016-0773-0>
- Lowe, T. W. (2015). Online quizzes for distance learning of mathematics. *Teaching Mathematics and its Applications*, 34(3), 138–148. <https://doi.org/10.1093/teamat/hrv009>
- Luz, Y. & Yerushalmy, M. (2019). Students' conceptions through the lens of a dynamic online geometry assessment platform. *Journal of Mathematical Behavior*, 54. <https://doi.org/10.1016/j.jmathb.2018.12.001>
- Maharaj, A. & Dlomo, T. (2018). Revelations from an online diagnostic arithmetic and

- algebra quiz for incoming students. *South African Journal of Science*, 114(5–6), 1–7. <https://doi.org/10.17159/sajs.2018/20170078>
- Martins, S. G. (2017). Weekly online quizzes to a mathematics course for engineering students. *Teaching Mathematics and its Applications*, 36(1), 56–63. <https://doi.org/10.1093/teamat/hrw011>
- Martins, S. G. (2018). A study of the application of weekly online quizzes in two courses of mathematics for engineering students - is it a fair and effective strategy to increase student learning? *International Journal of Innovation in Science and Mathematics Education*, 26(1), 46–59.
- Massing, T., Schwinning, N., Striwe, M., Hanck, C. & Goedicke, M. (2018). E-Assessment Using Variable-Content Exercises in Mathematical Statistics. *Journal of Statistics Education*, 26(3), 174–189. <https://doi.org/10.1080/10691898.2018.1518121>
- Monaghan, J., Trouche, L. & Borwein, J. M. (2016). *Tools and Mathematics* (Vol. 110). Berlin: Springer, Cham. <https://doi.org/10.1007/978-3-319-02396-0>
- Murchan, D. & Oldham, E. (2017). Exploring the role of computer-based assessment in diagnosing children's mathematical errors in primary education in Ireland. *Irish Educational Studies*, 36(4), 489–510. <https://doi.org/10.1080/03323315.2017.1393765>
- Murphy, R., Roschelle, J., Feng, M. & Mason, C. A. (2020). Investigating Efficacy, Moderators and Mediators for an Online Mathematics Homework Intervention. *Journal of Research on Educational Effectiveness*, 13(2), 235–270. <https://doi.org/10.1080/19345747.2019.1710885>
- Ocaña, A., Pulido, D., Gil, S. & Zuluaga, M. (2019). Changes in the performance of students of a mathematical thinking course through formative assessment with an online question bank. *Interdisciplinaria*, 36(1), 7–22.
- OCDE. (2015). *Students, Computers and Learning. Making the connection*. Paris: OECD Publishing.
- Olsher, S. (2019). Making Good Practice Common Using Computer-Aided Formative Assessment. En G. Aldon & J. Trgalová (Eds.), *Technology in Mathematics Teaching* (pp. 31–47). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-19741-4_2
- Olsher, S., Yerushalmy, M. & Chazan, D. (2016). How might the use of technology in formative assessment support changes in mathematics teaching? *For the Learning of Mathematics*, 36(3), 11–18.
- Pacheco-Venegas, N. D., López, G. & Andrade-Aréchiga, M. (2015). Conceptualization, development and implementation of a web-based system for automatic evaluation of mathematical expressions. *Computers & Education*, 88, 15–28.
- Pásztor, A., Molnár, G. & Csapó, B. (2015). Technology-based assessment of creativity in educational context: The case of divergent thinking and its relation to mathematical achievement. *Thinking Skills and Creativity*, 18, 32–42. <https://doi.org/10.1016/j.tsc.2015.05.004>
- Pelkola, T., Rasila, A. & Sangwin, C. (2018). Investigating bloom's learning for mastery in mathematics with online assessment. *Informatics in Education*, 17(2), 363–380. <https://doi.org/10.15388/infedu.2018.19>

- Rabardel, P. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains*. Paris: Armand Colin.
- Ramamurthy, M., Kousika, N., Mohana, G. S. & Mageswari, K. (2017). Scope of E-assessment of answers in mathematical domain. *International Journal of Civil Engineering and Technology*, 8(10), 533–537.
- Richard, P., Venant, F. & Gagnon, M. (2019). *Issues and Challenges in Instrumental Proof*. (D. Reid, Ed.). Cham: Springer International Publishing. https://doi.org/10.1007/978-981-32-9323-6_1
- Rønning, F. (2017). Influence of computer-aided assessment on ways of working with mathematics. *Teaching Mathematics and its Applications*, 36(2), 94–107. <https://doi.org/10.1093/teamat/hrx001>
- Roschelle, J., Feng, M., Murphy, R. F. & Mason, C. A. (2016). Online Mathematics Homework Increases Student Achievement. *AERA Open*, 2(4), 1–12. <https://doi.org/10.1177/2332858416673968>
- Sangwin, C. (2015). Inequalities, assessment and computer algebra. *International Journal of Mathematical Education in Science and Technology*, 46(1), 76–93. <https://doi.org/10.1080/0020739X.2014.941424>
- Sangwin, C. (2019). Reasoning by Equivalence: The Potential Contribution of an Automatic Proof Checker. En *Proof Technology in Mathematics Research and Teaching* (pp. 313–330). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-28483-1_15
- Sangwin, C., Cazes, C., Lee, A. & Wong, K. (2010). Micro-level automatic assessment supported by digital technologies. En C Hoyles & J.-B. Lagrange (Eds.), *Mathematics Education and Technology-Rethinking the Terrain* (pp. 227–250). Boston MA: Springer US. https://doi.org/10.1007/978-1-4419-0146-0_10
- Sangwin, C. & Jones, I. (2017). Asymmetry in student achievement on multiple-choice and constructed-response items in reversible mathematics processes. *Educational Studies in Mathematics*, 94(2), 205–222. <https://doi.org/10.1007/s10649-016-9725-4>
- Sarmiento, C. P. (2017). Student perceptions of online homework in mathematics of accounting and finance. *Advanced Science Letters*, 23(2), 1122–1125. <https://doi.org/10.1166/asl.2017.7513>
- Shotwell, M. & Apigian, C. H. (2015). Student performance and success factors in learning business statistics in online vs. on-ground classes using a web-based assessment platform. *Journal of Statistics Education*, 23(1). <https://doi.org/10.1080/10691898.2015.11889727>
- Sikurajapathi, I., Henderson, K. & Gwynllyw, R. (2020). Using e-assessment to address mathematical misconceptions in engineering students. *International Journal of Information and Education Technology*, 10(5), 356–361. <https://doi.org/10.18178/ijiet.2020.10.5.1389>
- Smolinsky, L., Marx, B. D., Olafsson, G. & Ma, Y. A. (2020). Computer-Based and Paper-and-Pencil Tests: A Study in Calculus for STEM Majors. *Journal of Educational Computing Research*. <https://doi.org/10.1177/0735633120930235>
- Smolinsky, L., Olafsson, G., Marx, B. D. & Wang, G. (2019). Online and Handwritten Homework in Calculus for STEM Majors. *Journal of Educational Computing Research*, 57(6), 1513–1533. <https://doi.org/10.1177/0735633118800808>

- Stacey, K. & Wiliam, D. (2013). Technology and assessment in mathematics. En M. A. K. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 721–751). New York: Springer. https://doi.org/10.1007/978-1-4614-4684-2_23
- Sureda-Negre, J. & Comas-Forgas, R. (2017). La revisión bibliográfica desde una perspectiva sistémica. En S. R. Pantoja & J. A. Rasco (Eds.), *Investigación cualitativa en Educación* (pp. 387–400). Buenos Aires: Miño y Dávila Editores.
- Vasko, M., Ritter, S. & Metzger, G. (2018). Online homework in engineering mathematics: can we narrow the performance gap? *International Journal of Engineering Pedagogy*, 8(1), 29–42.
- Walkington, C., Clinton, V. & Sparks, A. (2019). *The effect of language modification of mathematics story problems on problem-solving in online homework. Instructional Science* (Vol. 47). Springer Netherlands. <https://doi.org/10.1007/s11251-019-09481-6>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>

Autor

Jorge Gaona

Doctor y Magister en Didáctica de las Matemáticas de la Universidad de Paris, Francia.
Licenciado en Matemáticas de la Pontificia Universidad Católica de Valparaíso, Chile.
Profesor e investigador en la Escuela de Pedagogía en Matemáticas de la Universidad Academia de Humanismo Cristiano en Santiago y profesor del magíster en Didáctica de las Matemáticas de la Pontificia Universidad Católica de Valparaíso. Sus líneas de interés son Tecnología, específicamente en Evaluación en Línea y Espacio de Trabajo Matemático.
Email: jgaonap@docentes.academia.cl