

Estudio del desempeño en prácticas científicas de indagación y argumentación de alumnado de enseñanza secundaria mediante una aplicación de móvil

Mauricio Aguilera y Víctor López-Simó

Universidad Autónoma de Barcelona, Departamento de Didáctica de la Matemática y de las Ciencias Experimentales, España. E-mails: mauricio.aguilera@autonoma.cat , victor.lopez@uab.cat.

Resumen: La finalidad de esta investigación es caracterizar los niveles de desempeño en prácticas científicas de indagación y argumentación alcanzado por el alumnado que participa en un proyecto de pensamiento crítico sobre la confiabilidad de una App móvil comercial. Se analizaron 76 videos elaborados por estudiantes de 2º de ESO donde presentaron una pequeña investigación para recolectar pruebas sobre una App a su elección. A partir de la revisión bibliográfica de distintos instrumentos de análisis para determinar el desempeño en indagación y argumentación, se generó un sistema de categorías con 5 niveles de desempeño para cada práctica científica, que se usó para analizar los videos. Los resultados mostraron un nivel de desempeño variable, pero con el grueso de estudiantes situado en niveles simples y con una ligera relación entre el nivel en indagación y argumentación. Además, se observan pequeñas diferencias cuando los resultados se desagregan en las tres implementaciones llevadas a cabo con tres profesores distintos y también según la App utilizada por los estudiantes. Estos resultados apuntan al potencial de trabajar en el aula de secundaria algunos aspectos del pensamiento crítico como la fiabilidad, validez y verificación.

Palabras clave: prácticas científicas, indagación, argumentación, móvil apps, enseñanza secundaria.

Title: Study of student performance in scientific inquiry and argumentation practices through a mobile application.

Abstract: The purpose of this research is to characterize the levels of performance in scientific inquiry and argumentation practices achieved by the students participating in a critical thinking project regarding the reliability of a commercial mobile app. Seventy-six videos created by 2nd-grade ESO students were analyzed, in which they presented a small investigation to gather evidence about a mobile app of their choice. Based on the review of various analytical tools to determine performance in inquiry and argumentation, a category system with 5 performance levels for each scientific practice was generated and used to analyze the videos. The results showed a variable level of performance, with the majority of students falling into the lower levels, and a slight correlation between the levels of inquiry and argumentation. Additionally, small differences were observed when the results were disaggregated across the three

implementations carried out with three different teachers and also based on the app used by the students. These results point to the potential of working on certain aspects of critical thinking, such as reliability, validity, and verification, in the high school classroom.

Keywords: scientific practices, inquiry, argumentation, mobile apps, secondary education.

Introducción

El desarrollo del pensamiento crítico de los estudiantes a lo largo de su escolaridad ha sido por décadas una cuestión de gran relevancia dentro del ámbito de investigación e innovación en educación científica (Jiménez-Aleixandre, 2009; Oliveras y Sanmartí, 2009; Blanco-López et al., 2017). Si bien existen diversas definiciones y aproximaciones al pensamiento crítico y su relación con la educación científica, lo entenderemos como un conjunto de disposiciones afectivas y, habilidades cognitivas y metacognitivas que configuran la capacidad de los individuos para emitir juicios en un contexto y de forma razonada (Facione, 1990). Algunas de las tradiciones didácticas de nuestro ámbito, como la argumentación científica o el uso educativo de controversias socio-científicas han estado estrechamente ligadas a su desarrollo proponiéndose diferentes formas de promoverlo y evaluarlo (Blanco-López et al., 2017; Cebrián-Robles et al., 2018; Vila et al., 2023). Sin embargo, a medida que nos adentramos en la denominada era de la "postverdad", la necesaria capacitación de los estudiantes para discernir de forma crítica y reflexiva la calidad de la información que recibe ha tomado todavía más relevancia (Couso y Puig, 2021; Osborne et al., 2022).

Actualmente, los profesionales de la educación científica se enfrentan a retos sociales como la propagación de bulos pseudocientíficos y el apoyo a todo tipo de teorías alternativas. Para hacer frente a estos fenómenos, algunas iniciativas ciudadanas, gubernamentales o periodísticas en diferentes lugares del mundo han propuesto formas de evaluar y verificar la confiabilidad de las informaciones que se difunden, tales como los denominados "fact check", sistemas de verificación mediante pruebas de las afirmaciones difundidas a través de redes sociales o expresadas por políticos y periodistas (Graves, 2017).

Inspirándose en esta lógica de la verificación, un grupo de innovación docente de profesores de física y química desarrollaron un proyecto de aula para estudiantes de secundaria llamado "App Checkers" (López-Simó, 2021), en el cual se propuso a estudiantes de ESO que verifican la confiabilidad de una aplicación móvil comercial (App) descargable en celular, y que supuestamente midiera algo del exterior, ya fuera a través de la cámara del teléfono, de su micrófono o de cualquiera de sus sensores internos. El estudiante elegía libremente que App investigar, siempre que fuere de carácter gratuito y segura, que incluyera algún tipo de medición del exterior (fuera cierta o no), y que no infligiera ningún tipo de norma de convivencia y respeto a los demás. En la figura 1 se muestran algunos ejemplos de Apps elegidas por los estudiantes.

Una vez elegida la App, los estudiantes en grupos de trabajo pequeños diseñan y llevan a cabo una investigación para recoger datos y responder a

la pregunta: ¿Cómo de confiable es la App elegida? Como producto final, debían elaborar un pequeño video argumentativo con su veredicto, determinando qué grado de confiabilidad daban a su App, apoyándose en las pruebas recogidas. El video final, de esta forma, incorpora una dimensión tanto de indagación como de argumentación.

Bajo este contexto es que surge la presente investigación, que tiene como propósito caracterizar la dimensión indagativa y argumentativa de los videos realizados por los estudiantes de 2º de ESO en el marco del proyecto App Checkers, para así, profundizar en la comprensión del potencial didáctico que tiene un proyecto de verificación de Apps para involucrar a los estudiantes en estas prácticas científicas (Jiménez-Aleixandre, 2010; Jiménez-Liso, 2020).

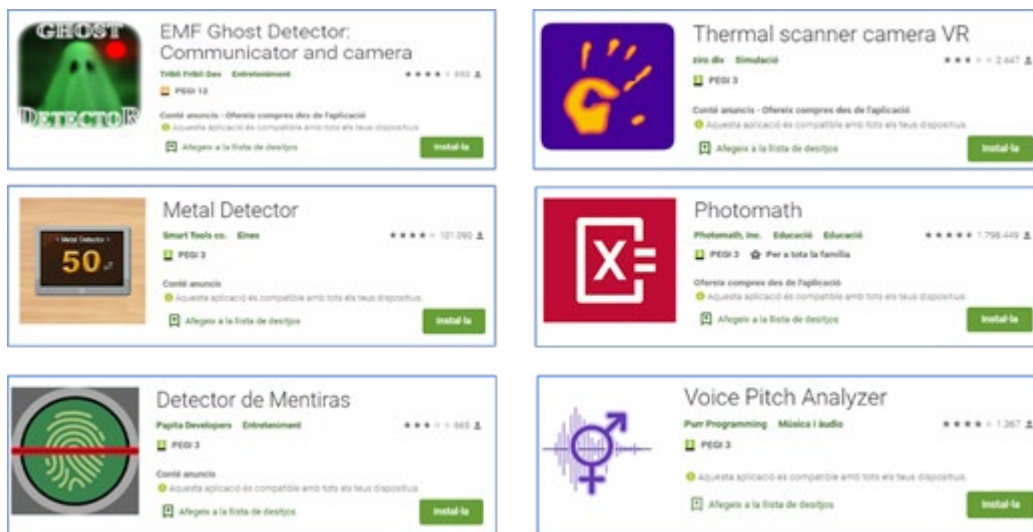


Figura 1. Seis ejemplos de Apps elegidas por los estudiantes en el proyecto App Checkers.

Marco teórico

Indagar científicamente sobre la confiabilidad de una App

El término indagación es un término polisémico, que suele ir asociado a diferentes aproximaciones según si se plantea como metodología o como objetivo de enseñanza (Couso, 2014). Para efectos de esta investigación, el interés estará referido a la capacidad de “indagar” o “investigar” científicamente que los estudiantes deben desarrollar. Para determinar esta capacidad han existido distintos tipos de instrumentos, como el denominado “Inventario de Evaluación de Pruebas Prácticas” (PTAI) diseñado por Tamir et al. (1982) para evaluar la capacidad investigativa de los estudiantes. Su creación tenía como objetivo describir un inventario de evaluación para exámenes prácticos de laboratorio orientados a la indagación. En las últimas décadas el PTAI ha sido la base con la que otros investigadores educativos han construido sus propios instrumentos de evaluación (Arnold et al., 2014; Franco-Mariscal, 2015). Una de estas adaptaciones fue realizada por Ferrés et al. (2015), que construyeron una rúbrica de evaluación de siglas NPTAI, del acrónimo Nuevo Inventario de Evaluación de Pruebas, con el fin de evaluar los procesos de una indagación más abierta y autónoma que el PTAI. El NPTAI también ha sido usado por diferentes investigadores, de

manera completa o parcial dependiendo del contexto y de la tarea solicitada a los estudiantes (Franco-Mariscal, 2015; Rosa, 2019; Pérez, 2019; Pozuelo y Cascarosa, 2018; Solé Llussà et al., 2017).

En paralelo a estas propuestas, Gott et al. (2003) también abordaron las capacidades indagativas de los estudiantes, señalando que existe un conjunto de conocimientos que subyacen a la comprensión de las pruebas científicas. Según estos autores, los Concepts to Evidence (CoE) refieren un conjunto de ideas que sustentan la recopilación, el análisis y la interpretación de los datos, y que deben entenderse por parte de los estudiantes antes de poder manejar las pruebas, esto quiere decir que los datos deben ser sometidos a algún tipo de validación para que sea posible asignarles un peso, luego de este proceso de validación y fiabilidad pueden ser usados como pruebas. Por ello las capacidades indagativas son tan importantes, ya que permiten evaluar la calidad del experimento y las condiciones en las que se llevó a cabo, junto con el análisis comparativo de lo reproducido por otros experimentadores en otras circunstancias y la practicidad de implementar sus resultados. Sin embargo, para otros investigadores como Bravo et al. (2009), el estudio de Gott et al. (2003) se refiere a trabajos prácticos de naturaleza empírica de los estudiantes y, por tanto, no es específica de una competencia de uso de pruebas que sería más global, pues falta incluir, por ejemplo, aquellas pruebas de naturaleza teórica. Sin embargo, Bravo et al. (2009) admiten que el citado trabajo es un marco de referencia completo dentro de lo que es específico para el uso de pruebas empíricas.

Según Gott et al. (2003) para que un diseño experimental se considere fiable y válido deben haberse cuestionado diferentes aspectos como, si los instrumentos son atingentes; si la exactitud y precisión de las mediciones son adecuadas para determinar patrones entre variables; el tamaño de la muestra y las técnicas de muestreo; si el rango de mediciones y número de lecturas son adecuadas; si las variables son correctas; y si los controles de variables y la prueba justa son válidos, entre otras cosas. Una buena indagación implica que cada dato obtenido del diseño requiere un promedio de varias lecturas repetidas, cuyo número depende de la precisión requerida según cada contexto. Además, puesto que los instrumentos pueden estar sujetos a inexactitudes inherentes, el uso de diferentes instrumentos debe contribuir a aumentar la fiabilidad y validez de la indagación. En este sentido, el error humano y del instrumento de medida debe ser disminuido mediante comprobaciones aleatorias e independientes de las técnicas de medición por parte de otros experimentadores o grupos de investigación. Finalmente, Gott et al. (2003) señalan la importancia de realizar varios experimentos; comparar con datos secundarios recopilados de otros investigadores o fuentes confiables; y llevar a cabo una triangulación de los datos.

Argumentar científicamente sobre la confiabilidad de una App

Igual que sucede en la indagación, también existen diferentes acepciones del término argumentación, aunque la más aceptada en nuestro ámbito se refiere a la evaluación del conocimiento en base a pruebas, tal como propone Jiménez-Aleixandre (2010). La argumentación científica requiere que el estudiante participe de actividades en la cuales sea necesario

movilice recursos de tipo cognitivo y metacognitivo, que incorpore diseños en contextos científicos puros o en contextos socio-científicos, en los que el estudiante debe tener en cuenta cuestiones éticas, sociales, y/o económicas (López-Gay et al., 2020).

Para el estudio de la argumentación científica, uno de los referentes más usados es el modelo de Toulmin, el cual propone que la estructura de una argumentación consta de seis componentes, cada una con una función determinada: pruebas (P), justificación (J), respaldo (B), aserción (A), calificadores modales (M) y refutación (R) (Toulmin, 2003). Se destaca que la refutación puede verse en dos sentidos: primero, la refutación a las pruebas de otro, esto en el marco de un diálogo entre dos o más interlocutores y, segundo, las condiciones de refutación, que son el reconocimiento de las restricciones o excepciones que se aplican a la aserción.

Diferentes referentes de la literatura didáctica han utilizado el modelo de Toulmin (2003), entre otras cosas, para crear dos tipos de instrumentos para evaluar el nivel de desempeño de los estudiantes en argumentación. El primero lo componen los sistemas de análisis que gradúan cada una de las seis componentes del modelo de Toulmin en niveles de desempeño. Así, por ejemplo, existen instrumentos de análisis de argumentación que constan con 4 niveles de desempeño para la componente pruebas, 4 niveles más para la categoría de justificación, y así con cada componente. Dentro de esta tipología, en algunos casos se asignan puntuaciones para obtener una valoración mediante la suma total, de manera que se puede evaluar por separado la calidad de cada componente (Cho y Jonassen, 2002; Cebrián-Robles et al., 2018; Yeh y She, 2010).

En paralelo, el segundo tipo de instrumentos integran una a una las componentes del modelo, de modo que un mayor número de componentes equivale a un mejor desempeño argumentativo (Lin et al., 2012; Osborne, 2005). En este tipo de instrumento no se pierde aparentemente la interrelación entre las componentes de una argumentación definidas en el modelo de Toulmin. Este último está en acuerdo con los desarrollos de progresiones de aprendizaje trabajados por Osborne et al. (2016) y Berland y McNeill (2010). Las herramientas usadas en este segundo tipo de análisis suelen considerar como premisa que el núcleo de toda argumentación es: prueba, justificación y aserción (Sardá y Sanmartí, 2000; Jiménez-Aleixandre, 2010), sin las cuales un argumento no sería válido o, dicho de otra forma, no calificaría como argumento. De esta manera, las otras componentes del modelo, garantía, calificadores modales y refutación, se pueden considerar auxiliares (Jiménez-Aleixandre, 2009). Lo anterior define un nivel de desempeño inferior que corresponde a aquellos productos argumentativos que no poseen este núcleo, luego en un nivel posterior están los que presentan este núcleo y finalmente los que ya integran en su argumentación otras componentes, siendo la refutación un nivel de desempeño superior, pues representa una mayor demanda cognitiva (Osborne et al., 2016). Esto es así porque una aserción debe ser defendida con pruebas para ser parte de un argumento ya que uno no está participando en la práctica de la argumentación científica si simplemente están haciendo valer la aserción (Berland y McNeill, 2010). En cuanto a la jerarquía, las componentes justificación y respaldo están diseñadas para

explicar cómo las pruebas apoyan la aserción, por lo tanto, uno debe tener pruebas antes de que pueda tener un razonamiento. Así, se sugiere que los productos argumentativos aumenten en calidad a medida que se agregan más componentes a un argumento, iniciando con las pruebas. Esta idea fue desarrollada por Osborne et al. (2016) en la construcción y validación empírica de su progresión de aprendizaje en argumentación.

A pesar del potencial que tiene el análisis en argumentación a través de las componentes del modelo de Toulmin, cabe destacar también algunas limitaciones metodológicas al usar el modelo. Existen problemas con respecto al "tamaño de grano" de la información que se busca y utiliza, y también sucede que muchas veces las categorías construidas para cada componente se solapan (Duschl, 2007; Díaz de Bustamante y Jiménez-Aleixandre, 1999). Así, una de las dificultades que se ha establecido, está en el esclarecimiento de lo que cuenta como prueba, justificación o respaldo. Algunos autores han integrado estas dos últimas en una sola categoría, con el fin de eludir algunos de estos problemas metodológicos de solapamiento (Berland y McNeill, 2010; McNeill y Krajcik, 2007; Sampson et al., 2013; Osborne et al., 2016; Erduran et al., 2004). También se visualiza que existe una diferencia entre prueba y dato, que es en general entendido más bien como una medición de carácter numérica o cualitativa, mientras que prueba es un concepto más general, si algo es prueba depende de si apoya la aserción, es decir, de su papel en la argumentación. Cuando el dato es empírico y se considera como prueba, su jerarquía dentro del argumento debería ser funcional respecto a la fiabilidad y validez con que ese dato fue obtenido (Gott et al., 2003). Que exista un proceso para que un dato se transforme en prueba, implica que existe un tipo de valor epistémico en ellas y que admiten algún tipo de clasificación. Según Tytler et al. (2001) un tipo de clasificación podría ser: (1) pruebas científicas formales basadas en datos, (2) pruebas informales, y (3) cuestiones más amplias que inciden en las pruebas (intereses medioambientales, legales o económicos, entre otros).

Contexto y objetivos de investigación

Tal como se ha planteado en la introducción, App Checkers (López-Simó, 2021) es una actividad que promueve la estrategia de aprendizaje basado en proyectos para verificar científicamente la confiabilidad de una App móvil comercial que supuestamente mida alguna magnitud (sea cierta o no) y que pueda tener aplicaciones sociales y/o profesionales. El proyecto fue concretado en 4 etapas claramente definidas.

1a Etapa - Discusión inicial sobre concepto de verificación: Trabajo de dos horas en torno a la pregunta: ¿Qué significa que una App que tiene por objetivo la medición, sea falsa o engañosa? Se realizó una introducción donde los estudiantes visualizaron vídeos de Fact Check para reflexionar sobre el significado de "verificación". Posteriormente se trabajó alrededor de sus ideas iniciales sobre "medir" y "confiabilidad", para luego explorar diferentes Apps, gratuitas y de contenido apto para menores de edad, indagaron sobre su funcionamiento y analizaron los comentarios y sus valoraciones para finalmente elaborar de un listado de Apps posibles de ser verificadas.

2ª Etapa - Selección de una App y diseño de una investigación: Trabajo de dos horas en torno a la pregunta: ¿Qué significa que una App que tiene por objetivo la medición, sea falsa o engañosa? Se realizó una introducción donde los estudiantes visualizaron vídeos de Fact Check para reflexionar sobre el significado de "verificación". Posteriormente se trabajó alrededor de sus ideas iniciales sobre "medir" y "confiabilidad", para luego explorar diferentes Apps, gratuitas y de contenido apto para menores de edad, indagaron sobre su funcionamiento y analizaron los comentarios y sus valoraciones para finalmente elaborar de un listado de Apps posibles de ser verificadas.

3ª Etapa - Definición de una escala de confiabilidad: Los estudiantes, durante 2 horas de clase, diseñaron y ejecutaron un experimento para poner a prueba una App. Las acciones que los estudiantes llevaron a cabo fueron la elección de la App, seguidamente de la planificación de uno o varios experimentos para poner a prueba la App y recoger resultados.

4ª Etapa - Elaboración de un video: Se dedicaron 3 horas de trabajo y consistió en la elaboración de un guion por parte de los estudiantes que posteriormente facilitó la elaboración y edición de una video-presentación con una duración de entre 2 y 4 minutos.

El proyecto fue realizado en el Instituto Pau Vila de Sabadell en Cataluña, un instituto público con aproximadamente 150 estudiantes por curso organizado en 5 grupos-clases heterogéneos. Se trabajó con estudiantes de segundo año de ESO durante dos años académicos distintos, con tres profesores distintos como se muestra en la tabla 1.

Profesor	Año académico	Nº videos
Profesor 1	2019-20	29
Profesor 2	2020-21	35
Profesor 3	2020-21	12

Tabla 1. Profesores que implementaron App Checkers.

Durante el proyecto se observó que la variedad de Apps elegidas por los estudiantes era muy grande, por lo que surgió la pregunta de si el tipo de App había influido en el desempeño en indagación y argumentación que mostraban los estudiantes.

Además, el hecho de implementar el proyecto durante 3 cursos distintos conllevó algunas variaciones internas. En el año académico 19-20 se implementó el proyecto con el profesor 1 que había diseñado el proyecto, y en el curso 20-21 se volvió a implementar con otros dos profesores distintos (profesor 1 y profesor 2), que, por limitaciones de tiempo, decidieron prescindir de la actividad 3 del proyecto. Esta corresponde a la "Definición de una escala de confiabilidad" de la figura 2, que había servido como andamiaje para que los estudiantes del curso 19-20 tuviera criterios para valorar la confiabilidad de la App.

Dada la variedad de videos elaborados por los estudiantes y el posible efecto tanto del andamiaje usado como del tipo de App elegida, nos planteamos como objetivo de investigación caracterizar el nivel de

desempeño en las prácticas científicas de indagación y argumentación de los estudiantes participantes en el proyecto App Checkers.

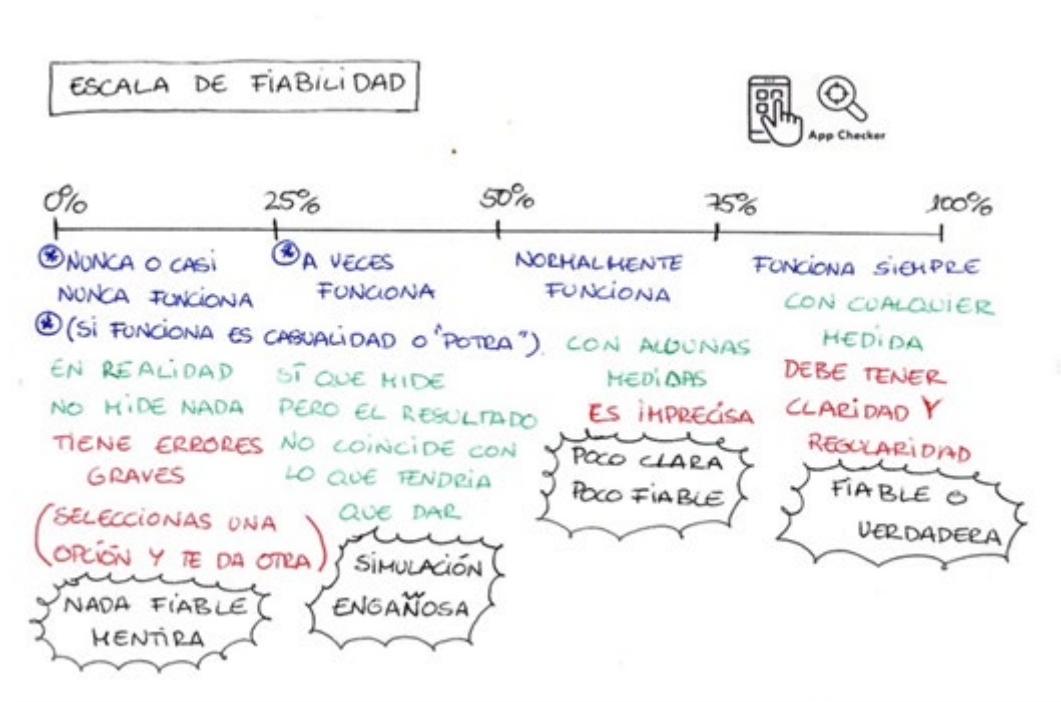


Figura 2. Fotografía del dossier de un estudiante donde construye una escala de confiabilidad entre 0% y 100%. Esta actividad se usó en el curso 2019-20, pero no en el 2020-21.

Metodología

Recogida de datos

En las distintas implementaciones del proyecto App Checkers el producto final indagativo y argumentativo ha sido un video, el cual fue realizado en sus casas, flexibilidad otorgada inicialmente a causa del confinamiento por la pandemia del COVID-19. Del curso 19-20 se recibieron un total de 29 videos, de los cuales 22 eran individuales, 5 en parejas, 1 en trío y 1 de cuatro estudiantes. De las cohortes 20-21 se recibieron 47 videos de los cuales 20 fueron individuales, 14 en parejas, 9 en tríos, 3 de a cuatro y 1 de a cinco estudiantes. En la tabla 2 se muestra la mayoría de Apps con las cuales trabajaron los estudiantes.

El procedimiento de tratamiento de datos correspondió primero a una codificación con la finalidad de sistematizar los videos en una base de datos y luego una transcripción de cada video con la finalidad de estudiar en profundidad los videos de los estudiantes y el trabajo en indagación y argumentación.

Categorías para el análisis del desempeño en indagación

Utilizamos el PTAI (Tamir et al., 1982), NPTAI (Ferrés et al., 2015) y los CoE (Gott et al., 2003) para extraer indicadores que nos permitieran determinar categorías de desempeños respecto de las acciones que podrían realizar los estudiantes. Las categorías para determinar niveles de

desempeño en indagación quedaron determinadas por cinco niveles que presentamos en el Anexo.

Tipo de App	Número de videos			
	Curso 2019-20 Profesor 1	Curso 2020-21 Profesor 2	Curso 2020-21 Profesor 3	Total
Detector de Mentiras	11	3	0	14
Detector de Metales	3	5	0	8
Cámara de Rayos X	4	1	2	7
Medidor de longitudes	0	7	0	7
Sonómetro	2	5	0	7
Detector de fantasmas	2	2	2	6
Medidor de temperatura	0	2	1	3
Detector de edad	1	2	0	3
Detector Infrarrojo	1	0	1	2
Buscador de celular	0	1	1	2
Medidor de luz	0	2	0	2
Otros	5	5	5	15

Tabla 2. Número de videos y Tipo de App para cada implementación. La categoría otros corresponden a Apps elegidas que sólo aparecen una vez.

Categorías para el análisis del desempeño en argumentación

Utilizamos los trabajos de Jiménez-Aleixandre (2010), Sarda y Sanmartí (2000), Berland y McNeill (2010), Sampson et al. (2013) y Osborne et al. (2016) para determinar niveles de desempeño en argumentación, donde queda establecido que la incorporación de más componentes del modelo de Toulmin implica mayores niveles de desempeño. Unimos en una sola categoría justificación y respaldo, la que llamamos razonamiento, ya que App Checkers trabaja con argumentos multimodales y nos permite manejar la amplia variedad de Apps recibida. También se estableció trabajar la componente refutación como "condiciones de refutación", que son el reconocimiento de las restricciones o excepciones que se aplican a la aserción (Jiménez-Aleixandre, 2010). Utilizamos el hecho que la triada Prueba-Razonamiento-Aserción implica un mínimo para argumentar con el cual establecemos el nivel de argumentación más simple. Las categorías para determinar niveles de desempeño en argumentación quedaron determinadas por cinco niveles de desempeño que presentamos en el Anexo.

Resultados

A partir de la categorización para los niveles de desempeño en argumentación e indagación, obtuvimos el gráfico de la figura 3, en el cual se presenta la cantidad de videos situados en cada nivel, mostrando los niveles de indagación en el eje horizontal y los niveles de argumentación en el eje vertical. Estos niveles se han denotado como I0, I1, I2, I3 e I4 para indagación y como A0, A1, A2, A3 y A4 para argumentación. El recuento de videos que caen en una intersección de niveles se representa con un círculo cuyo diámetro se corresponde con el número de videos en cada coordenada.

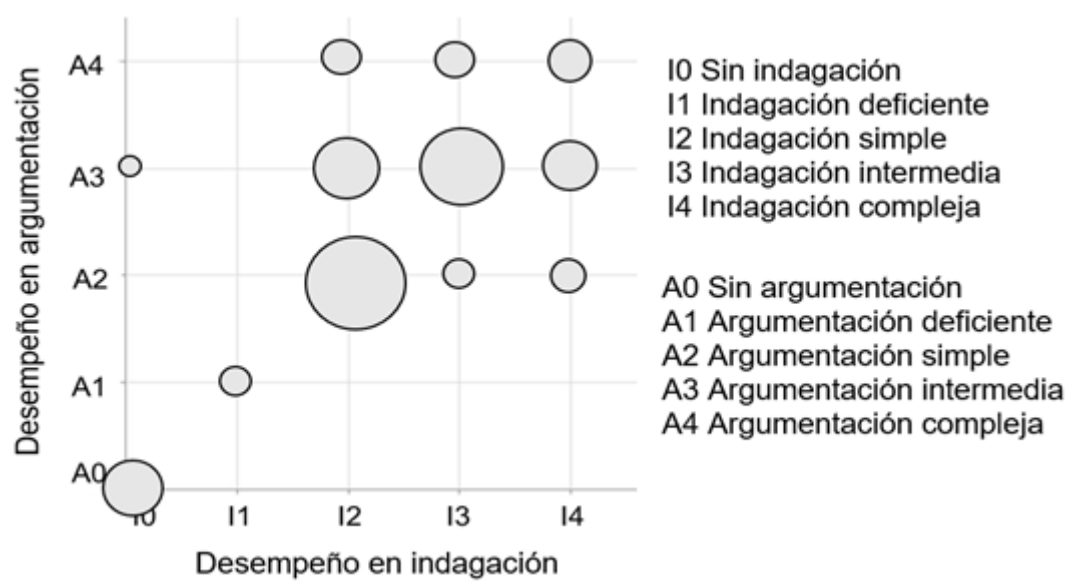


Figura 3. Distribución de los niveles de desempeño para los n = 76 videos.

En la figura 3 se aprecia que frente a la misma pregunta investigativa encontramos una amplia variedad de desempeño indagativo y argumentativo, pero con algunas regularidades. Podemos observar que la mayoría de los videos se sitúan en el cuadrante superior - derecho del gráfico, en los rangos entre I2 e I4 y, A2 y A4, aunque con menor presencia en los niveles I4 y A4 que en los niveles inferiores, coincidiendo así con lo expuesto en investigaciones previas (Osborne et al. 2016; Berland y McNeill, 2010; Yeah y She, 2010; Hsu et al., 2015). Además, del mismo gráfico se puede observar que las combinaciones más concurridas son aquellas que se sitúan en el eje diagonal del gráfico en las coordenadas I0-A0, I2-A2 y I3-A3, por lo que parecería que una mejor indagación coincide con una mejor argumentación.

Así, en primer lugar, los videos categorizados en el nivel de coordenadas I0-A0 (n=7) son aquellos donde no se ha planteado una investigación que permita obtener pruebas, por lo que tampoco existe argumentación alguna, tal como se ejemplifica en la siguiente transcripción:

“He elegido la App medidor de distancias. Puedes medir distancia corta o larga [Muestra las opciones por pantalla, enfoca una estantería de su habitación y usa la App para medir su longitud]. El resultado es 160,56 m [El resultado no tiene sentido, ya que la estantería se encuentra a unos 2 metros (Figura 4)]. Lo bueno que tiene la App es que el resultado se queda guardado. Esta es mi App, espero que os haya gustado.”



Figura 4. Capturas para un video situado en la coordenada I0-A0.

En segundo lugar, los videos de nivel I2-A2, que son los más numerosos (n=20), muestran un nivel de indagación y argumentación simples, ya que los estudiantes sólo toman una o dos lecturas válidas en cuanto al uso instrumental, y además no incluyen ningún tipo de refutación ni modulación. Un ejemplo es la App "PlantNet" de la figura 5 que, dada una imagen, determina qué tipo de planta contiene esa imagen. En este ejemplo, un grupo de estudiantes trabajó la App tomando dos mediciones, y comprobaron que las imágenes de salida y entrada eran similares, dando así por válida la App sin aportar mayor matiz. La transcripción es la siguiente:

"Verificaremos la App PlantNet, detecta plantas mediante la cámara haciendo una foto [*Explican las opciones y como operar la App*]. Comenzaremos con un cactus [*toman una foto*] (Figura 5), y aquí pone fotos de cactus, o sea que sí reconoce que planta es. Ahora tomaremos una foto a esta planta que es un Aloe Vera [*toman la foto*] (Figura 5), y aparece un Aloe Vera, si la reconoce. Hemos verificado que esta aplicación es cierta, porque salen fotos de la misma planta que hemos hecho la foto, ha funcionado con todas las plantas mostradas."

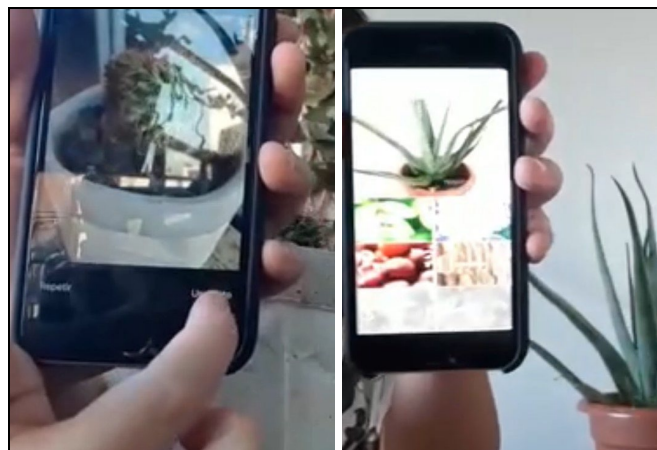


Figura 5. Transcripción y capturas de un video situado en la coordenada I2-A2.

La segunda combinación de niveles más frecuente es I3-A3 (n=15), que corresponde a una indagación y argumentación intermedia. Un ejemplo de

videos en este grupo corresponde a la figura 6, donde un estudiante usó la App "Thermal Scanner" que simula un escáner térmico, llevó a cabo tres experimentos diferentes recogiendo varias mediciones y concluyendo que la App cambiaba los colores entre imágenes de forma aleatoria. La transcripción es la siguiente:

"Utilizaremos la App Thermal Scanner. Es una aplicación de cámara térmica [luego, explica cómo funciona la App y sus opciones]. Para este experimento utilizaremos agua fría en vasos de colores y vasos transparentes y también utilizaremos una tostadora. Para empezar, en los dos vasos transparentes he puesto agua fría y agua caliente, y pone que están a la misma temperatura. Ahora en estos tres vasos de colores me he puesto agua fría a igual temperatura (Figura 6), y sale que uno está muy frío y los demás están calientes, además se ve cómo cambia los colores. Por tanto, esta aplicación sólo cambia los colores. Y para terminar utilizaremos la tostadora. Ahora está fría y después la encenderemos, para ver si cambia o no. Vamos a ver [la enciende y muestra la imagen de la App], pone que las partes que están calientes están frías (Figura 6). Las líneas estas azules indican frío, pero están calientes. Después de realizar este experimento llegamos a la conclusión que la App es falsa y sólo cambia los colores de los objetos por otros."

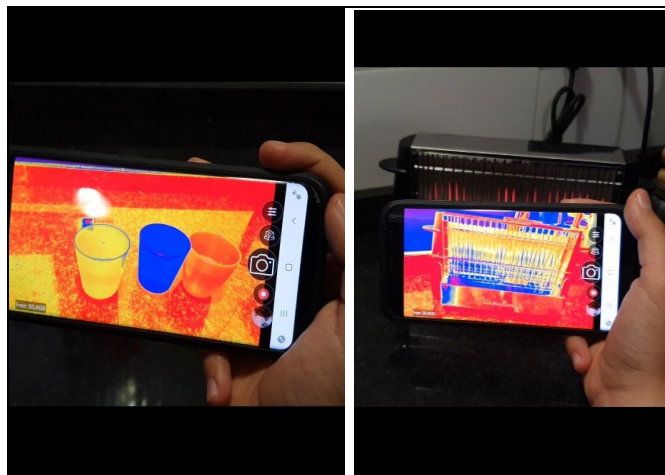


Figura 6. Transcripción y capturas de un video situado en la coordenada I3-A3.

Finalmente, dentro de la diagonal del gráfico de la figura 3 cabe destacar los videos situados en la coordenada I4-A4 (n=4), donde encontramos simultáneamente una indagación y argumentación complejas. En la figura 7 se muestra un ejemplo que corresponde al trabajo de un estudiante que indagó sobre la App "Photomath", puso a prueba la App de distintas maneras, usando tablas para obtener gráficos, y explorando sus alcances en cuanto a reconocimiento de imágenes entregándole diferentes tipos, y finalmente asignando una confiabilidad del 75%, presentando así una modulación en su argumentación y explicando estos matices en términos de refutaciones sobre la capacidad de la App en reconocer imágenes. La transcripción es la siguiente:

"Usamos la App "Photomath", que resuelve cuestiones matemáticas. Primero plantearemos una ecuación escrita a mano para ver si sabe detectarla y resolver (Figura 7), pasa la prueba. Ahora le pediremos que

resuelva un sistema de ecuaciones. También enfocaremos el número del ejercicio, y la palabra solución. Lo hace perfectamente y, además, nos deja elegir el método de resolución y muestra todos los pasos hasta la solución. Quizás uno de los mayores problemas está en el enfoque de la cámara. Aquí se puede ver cómo el escaneo de la imagen falla. Ahora intentaremos que nos represente en forma de gráfica las siguientes tablas [usan dos]. No las detecta. La siguiente prueba consistirá dar las fórmulas de las tablas. En ese caso sí lo resuelve en forma de gráfica. Por último, intentaremos que resuelva un problema de enunciado y algo sin sentido, y no sabe detectarlo (Figura 7). Teniendo en cuenta que le hemos puesto retos muy difíciles, creo que sería justo darle una confiabilidad del 75%.”

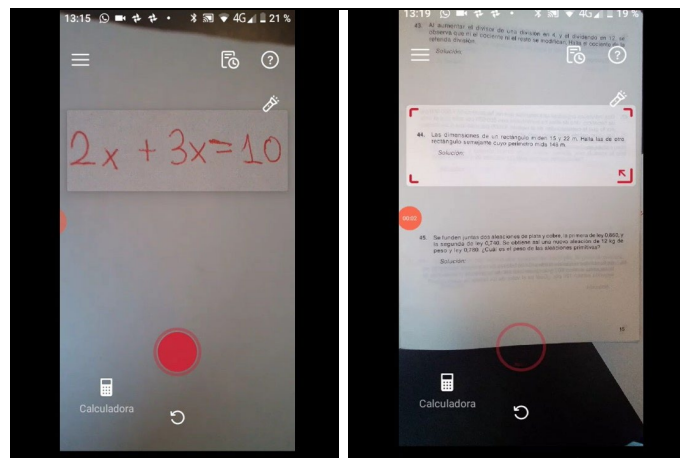


Figura 7. Transcripción y capturas de un video situado en la coordenada I4-A4.

A pesar de que el grueso de los videos se encuentre en la diagonal del gráfico, esto no implica que un mayor desempeño en indagación siempre vaya de la mano de un mayor desempeño en argumentación. Por un lado, encontramos algunos pocos videos en las coordenadas I3-A2 (n=2) e I4-A2 (n=3), que corresponde a desempeños indagativos altos que aportan pruebas robustas que luego no usan para argumentar adecuadamente, y también en la coordenada I4-A3 (n=6), donde a pesar de presentar una indagación compleja solamente desarrollan una argumentación intermedia.

Por otro lado, los videos que se sitúan por encima de la diagonal del gráfico se corresponden con indagaciones más simples y argumentaciones más elaboradas, incluyendo las coordenadas I2-A3 (n=10), I2-A4 (n=3) e I3-A4 (n=3). En todos estos casos observados, los datos que obtienen los estudiantes en su investigación son simples, pero los usan adecuadamente en su argumentación e incluyen otros datos informales que le permite modular su argumento o refutarlo parcialmente. Por ejemplo, en uno de los videos de la coordenada I2-A4 vemos un estudiante usando la App “Sonómetro” para tomar solamente dos medidas de sonido del televisor (con el volumen alto y bajo respectivamente), y mostrando cómo la aguja que medía los dB de la App subía y bajaba. A pesar de la simplicidad del experimento (indagación de nivel I2), donde no toma un número de mediciones suficientemente amplio que permita asegurar el patrón presentado como prueba, posteriormente señala que en ausencia de sonido la App no marca 0 dB tal como él había previsto, por lo que añade matices

al grado de confiabilidad que le ha dado a la App, en forma de modulación y refutación. Otro ejemplo, quizás el que presenta mayor discrepancia entre el nivel de indagación y argumentación, es el situado en la coordenada I0-A3 (n=1), donde un estudiante no investiga científicamente la App, pero argumenta basándose en una prueba informal como son los comentarios y la valoración de la App dados por otros usuarios, y presentando estas valoraciones para construir su argumento.

En paralelo, si estos resultados (n=76) se desagregan según las tres implementaciones hechas, una en el 2019-20 (n=29) y dos en el 2020-21 (n=35 y n=12), se obtienen los tres gráficos de la figura 8, donde se observan pequeñas diferencias en la distribución de niveles de desempeño.

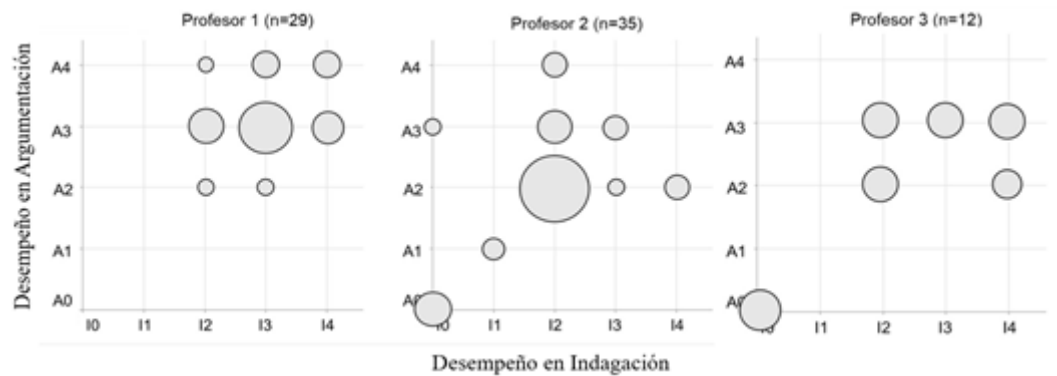


Figura 8. Distribución de los niveles de desempeño para indagación y argumentación según implementación.

El gráfico de la izquierda corresponde a los 29 videos de los estudiantes de la primera implementación, donde el profesor 1 usó una escala de confiabilidad (figura 2) para orientar a los estudiantes en su veredicto respecto a su App, las fases 1 a 3 fueron realizadas en el aula y los videos en sus casas. En este primer gráfico se observa que 11 de 29 videos (38%) se sitúan en la coordenada I3-A3 y que en 29 de 29 (100%) se puede identificar la triada Prueba – Razonamiento – Aserción (nivel de indagación y argumentación ≥ 2). En cambio, el gráfico del centro corresponde a 35 videos con el profesor 2, que no usó en su proyecto la escala de confiabilidad y además abordó el proyecto de una forma más superficial, por lo que muchos estudiantes se limitaron a demostrar si la App funcionaba o no funcionaba. Se observa que 17 de 35 videos (49%) se sitúan en la coordenada I2-A2. Finalmente, en el gráfico de la derecha se muestra el análisis de los 12 videos recogidos por el profesor 3 que tampoco usó la escala de confiabilidad y planteó el proyecto como un contexto para trabajar el método científico. Se observa que tanto con el profesor 2 como con el profesor 3 algunos videos se sitúan en la coordenada I0-A0 mostrada en la figura 4, y que corresponde a los estudiantes que no llegaron a entender la finalidad del proyecto. Todas estas diferencias en el nivel de desempeño del alumnado según la implementación realizada por cada profesor (poniendo más o menos énfasis en la escala de confiabilidad, en el diseño experimental, etc) sugieren una cierta influencia didáctica en el producto final, pero que debido a las limitaciones metodológicas no permiten establecer una única relación causal, sino simplemente una tendencia a explorar en futuras implementaciones del proyecto.

Finalmente, al desagregar los datos correspondientes a los 76 videos según el tipo de App elegida, se obtienen los gráficos de la figura 9, donde se presentan las 6 Apps más usadas por los estudiantes de acuerdo con la tabla 2. En estos gráficos se muestra que una misma App lleva a los estudiantes a diferentes niveles de desempeño tanto en indagación como en argumentación. Si comparamos los videos producidos por diferentes estudiantes que analizan una misma App, todos se distribuyen como mínimo en 3 niveles distintos de indagación y/o de argumentación, de manera que la variedad en el nivel de desempeño dentro de cada conjunto de videos según la App elegida es mayor que las diferencias que pudiera haber entre diferentes Apps.

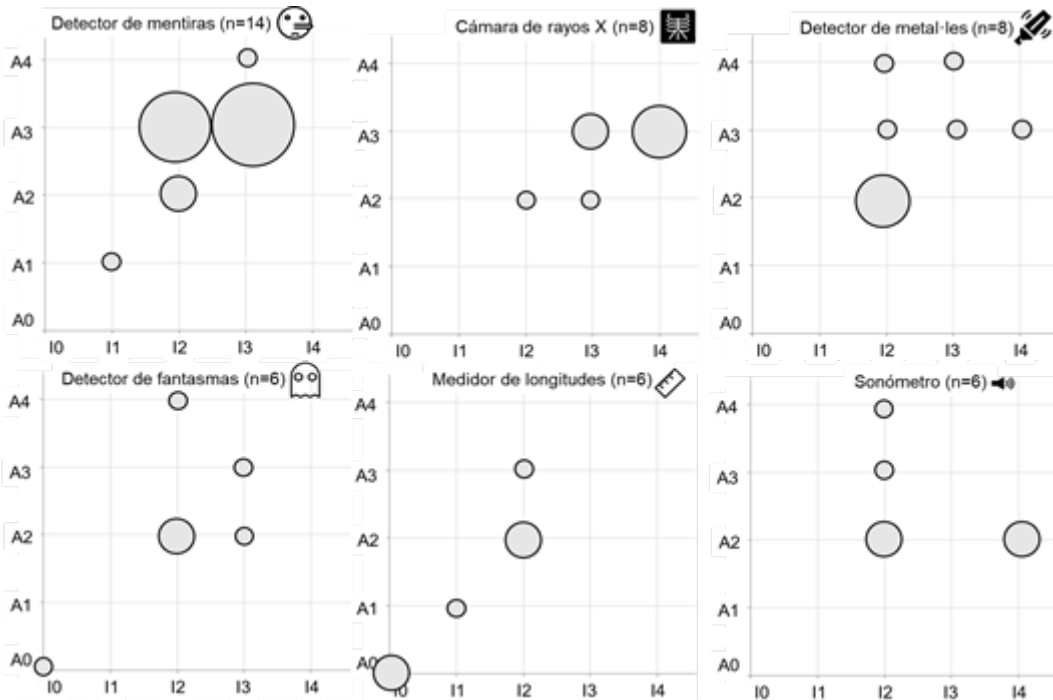


Figura 9. Distribución de los niveles de desempeño para indagación y argumentación según Tipo de App.

Las 3 Apps con una apariencia más científica (Rayos X, Detector de Metales y Sonómetro) son las únicas donde algunos estudiantes logran el nivel 4 de indagación, mientras que en las otras tres (Mentiras, Fantasmas y Longitudes) ningún grupo de estudiantes logra plantear indagaciones complejas. Del mismo modo, parecería que las Apps con una confiabilidad a priori más clara son en las que encontramos argumentaciones más sofisticadas, mientras que las Apps más confusas ningún estudiante logra un nivel de argumentación compleja. En el caso de Mentiras y Fantasmas todos los estudiantes creen que serán falsas y Metales y Sonómetro suelen recibir la confianza de los estudiantes desde el principio. Esto podría sugerir que los estudiantes argumentan mejor (añadiendo modulación y refutación) si se enfrenta a Apps cuya confiabilidad conocen de antemano, cosa que coincide con lo descrito por Osborne et al. (2016), donde el desempeño en argumentación es mayor en contextos donde los estudiantes se sienten más cómodos y con menor complejidad conceptual (Osborne et al., 2016). No obstante, debido al limitado tamaño de la muestra analizada, no se puede

generalizar esta tendencia, y sería necesario un análisis más detallado y sistemático.

Conclusiones

En esta investigación se han definido diferentes niveles de desempeño en las prácticas científicas de indagación y argumentación, con el objetivo de aplicarlos al análisis de los videos elaborados por estudiantes de 2º de ESO en el marco de un proyecto de verificación de Apps, llamado "App Checkers". Para ello, se ha partido de una revisión y adaptación de los diferentes instrumentos que permiten evaluar el grado de desempeño en indagación (Tamir et al.,1982; Ferrés et al., 2015; Gott et al., 2003) y en argumentación (Jiménez-Aleixandre, 2010; Sarda y Sanmartí, 2000; Berland y McNeill, 2010; Sampson et al.,2013; Osborne et al.,2016).

Así, cada video resultante del proyecto se ha analizado simultáneamente como producto argumentativo e indagativo, elaborando dos categorizaciones. Se obtiene una amplia variedad de niveles de desempeño en ambas dimensiones, debido a que las diferentes componentes que requiere una buena indagación y argumentación no aparecen de igual manera en todos los videos, sino que hay algunas que están mucho más presentes que otras, debido a las propias diferencias individuales de los estudiantes (Castellanos et al., 2002; Koerber y Osterhaus, 2019). Se han identificado algunas carencias en las habilidades de indagación, como, por ejemplo, la ausencia de calibración previa de algunas Apps que así lo requerían, o la ausencia de noción de este concepto y la habilidad para hacerlo (Gott et al. 2003).

También se ha identificado una tendencia según la cual los niveles de desempeño más altos en indagación corresponden en gran medida con niveles más altos de argumentación, pero no se puede hablar de una relación muy fuerte entre ambas dimensiones, ya que hay estudiantes que a pesar de plantear una indagación alta no usan los datos obtenidos para proponer una argumentación también alta, y en otros casos, a pesar de disponer de pocos datos (resultado de una indagación simple) pueden construir argumentos complejos (Osborne et al., 2016; Osborne, 2005). También se encuentran ligeras diferencias tanto según las diferentes implementaciones llevadas a cabo por diferentes profesores como según la App elegida, cuyas relaciones pueden ser el objeto de análisis de futuras investigaciones.

Así, la variedad de desempeños en indagación y argumentación identificadas en este proyecto señala algunas oportunidades importantes de enseñanza y aprendizaje, como son las ideas de verificación, de fiabilidad y validez, del concepto de medición y de error humano en la medición, de réplica, de refutación, etc. Sin embargo, dado que un grueso importante de estudiantes se queda en un nivel de argumentación e indagación simple o inferior, se pone de manifiesto la necesidad de promover las habilidades indagativas de los estudiantes para trabajar pruebas científicas y evaluar su confiabilidad, especialmente en un momento donde la enseñanza de las ciencias se enfrenta al reto de la desinformación (Osborne et al., 2022). Para ello, será necesario seguir investigando sobre qué estrategias educativas son más efectivas para dicho propósito, centrándose no

solamente en los productos finales que presentan los estudiantes al finalizar su proyecto de verificación, sino también adentrarse en los procesos y las prácticas en que participan los mismos (Jiménez-Aleixandre y Crujeiras-Pérez, 2017).

Agradecimientos

Esta investigación se ha financiado por el Ministerio de Ciencia e Innovación (PID2022-138166NB-C22b) y se ha realizado dentro del grupo de investigación SGR ACELEC, ref.2021 SGR 00647.

Referencias

Arnold, J., Kremer, K. y Mayer, J. (2014). Understanding Students' Experiments—What kind of support do they need in inquiry tasks? *International Journal of Science Education*, 36(16), 2719-2749. <https://doi.org/10.1080/09500693.2014.930209>

Berland, L., y McNeill, K. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, 94(5), 765-793. <https://doi.org/10.1002/sce.20402>

Blanco, A., España, E., y Franco-Mariscal, A. J. (2017). Estrategias didácticas para el desarrollo del pensamiento crítico en el aula de ciencias. *Ápice, Revista de Educación Científica*, 1(1), 107-115. <https://doi.org/10.17979/arec.2017.1.1.2004>

Bravo, B., Puig, B. y Jiménez-Aleixandre, M. P. (2009). Competencias en el uso de pruebas en argumentación. *Educación química*, 20(2), 137-142. [https://doi.org/10.1016/s0187-893x\(18\)30020-x](https://doi.org/10.1016/s0187-893x(18)30020-x)

Castellanos, D., Castellanos, B., Llivina, M., Silverio, M., Reinoso, C., y García, C. (2002). *Aprender y enseñar en la escuela*. La Habana, Cuba: Editorial Pueblo y Educación.

Cebrián-Robles, D., Cebrián de la Serna, M., Gallego-Arrufat, M. J., y Contreras, J. Q. (2018). Impacto de una rúbrica electrónica de argumentación científica en la metodología blended-learning. *RIED, Revista Iberoamericana de Educación a Distancia*, 21(1), 75-94. <http://dx.doi.org/10.5944/ried.21.1.18827>

Cho, K. L., y Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, 50(3), 5. <https://doi.org/10.1007/bf02505022>

Couso, D. (2014) De la moda de "aprender indagando" a la indagación para modelizar: una reflexión crítica. En M. Héras, A. Lorca, B. Vázquez, A. Wamba, R. Jiménez (eds). *Investigación y transferencia para una educación en ciencias: Un reto emocionante* (pp.1-28). Huelva: Servicio de Publicaciones Universidad de Huelva.

Couso, D., y Puig, B. (2021). Educación Científica en Tiempos de Pandemia. *Alambique: Didáctica de las ciencias experimentales*, 104, 49-56.

Díaz De Bustamante, J., y Jiménez-Aleixandre, M. P. (1999). Aprender ciencias, hacer ciencias: resolver problemas en clase. *Alambique. Didáctica de las ciencias experimentales*, 20, 9-16.

Duschl, R. A. (2007). Quality argumentation and epistemic criteria. En S. Erduran, M. P. Jiménez-Aleixandre (eds.). *Argumentation in science education: Perspectives from classroom-based research* (pp.159-175). Springer.

Erduran, S., Simon, S., y Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science education*, 88(6), 915-933. <https://doi.org/10.1002/sce.20012>

Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Newark. *American Philosophical Association*.

Franco-Mariscal, A. J. (2015). Competencias científicas en la enseñanza y el aprendizaje por investigación. Un estudio de caso sobre corrosión de metales en secundaria. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 33(2), 231-252. <https://doi.org/10.5565/rev/ensciencias.1645>

Ferrés, C., Marbà, A., y Sanmartí, N. (2015). Trabajos de indagación de los alumnos: instrumentos de evaluación e identificación de dificultades. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 12(1), 22-37. http://dx.doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2015.v12.i1.03

Graves, L. (2017) Anatomy of a Fact Check: Objective Practice and the Contested Epistemology of Fact Checking. *Communication, Culture and Critique*, 10(3) 518–537. <https://doi.org/10.1111/cccr.12163>

Gott, R., Duggan, S., Roberts, R., y Hussain, A. (2003). *Research into understanding scientific evidence*. <https://cofev.webspace.durham.ac.uk/>

Hsu, C. C., Chiu, C. H., Lin, C. H., y Wang, T. I. (2015). Enhancing skill in constructing scientific explanations using a structured argumentation scaffold in scientific inquiry. *Computers & Education*, 91, 46-59. <https://doi.org/10.1016/j.compedu.2015.09.009>

Jiménez-Aleixandre, M. P., Gallástegui, J. R., Eirexas, F., y Puig, B. (2009). *Actividades para trabajar el uso de pruebas y la argumentación en ciencias*. Danú

Jiménez-Aleixandre, M. P. (2010). *10 ideas clave. Competencias en argumentación y uso de pruebas*. Barcelona: GRAO.

Jiménez-Aleixandre M. P., y Crujeiras, B. (2017). Epistemic Practices and Scientific Practices in Science Education. En K. S. Taber y B. Akpan (Eds.), *Science Education. New Directions in Mathematics and Science Education* (pp. 69-80). Rotterdam: Sense Publishers.

Jiménez-Liso, M. R. (2020). Aprender ciencia escolar implica aprender a buscar pruebas para construir conocimiento (indagación). En D. Couso, M.

R. Jiménez-Liso, C. Refojo, J. A. Sacristán (coords.). *Enseñando Ciencia con Ciencia*. FECYT & Fundación Lilly. Madrid: Penguin Random House.

Koerber, S., y Osterhaus, C. (2019). Individual differences in early scientific thinking: Assessment, cognitive influences, and their relevance for science learning. *Journal of Cognition and Development, 20*(4), 510-533.

Lin, H., Hong, Z. R., y Lawrenz, F. (2012). Promoting and scaffolding argumentation through reflective asynchronous discussions. *Computers & Education, 59*(2), 378-384.

Llewellyn, D. (2012). *Teaching high school science through inquiry and argumentation*. Corwin Press

López-Gay, R., Jiménez-Liso, M. R., Martínez-Chico, M., y Castillo, F. J. (2020). Evidencias para la mejora de la enseñanza de las ciencias. *Dossier Graó, 5*, 39-43

López-Simó, V. (2021). App Checkers, un proyecto de verificación de la fiabilidad de una aplicación móvil. *Aula de secundaria, 44*, 37-42.

McNeill, K., y Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. En M. C. Lovett, P. Shah (eds.). *Thinking with data* (pp. 233-265). Lawrence Erlbaum Associates Publishers.

Oliveras, B., y Sanmartí, N. (2009). La lectura como medio para desarrollar el pensamiento crítico. *Educación química, 20*, 233-245. [https://doi.org/10.1016/s0187-893x\(18\)30058-2](https://doi.org/10.1016/s0187-893x(18)30058-2)

Osborne, J. (2005). The role of argumentation in science education. En K. Boersma (ed.). *Research and the quality of science education* (pp. 367-380). Amsterdam: Springer

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., y Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching, 53*(6), 821-846. <https://doi.org/10.1002/tea.21316>

Osborne, J., Pimentel, D., Alberts, B., Allchin, D., Barzilai, S., Bergstrom, C., Coffey, J., Donovan, B., Kivinen, K., Kozyreva, A., y Wineburg, S. (2022). *Science Education in an Age of Misinformation*. Stanford University: Stanford, CA.

Pérez, F. (2019). Obstáculos del Aprendizaje Basado en Problemas. Una experiencia pedagógica en el área de Bioquímica. *Voces y silencios, Revista Latinoamericana de Educación, 10*(2), 80-97.

Pozuelo, J., y Cascarosa, E. (2018). Inmersión en el mundo de la nanociencia a través de una experiencia de indagación guiada con alumnos de Educación Secundaria. *ReiDoCrea: Revista electrónica de investigación y docencia creativa, 7*, 376-387. <https://doi.org/10.30827/digibug.54141>

Rosa, S. (2019). Proyectos de investigación en los estudios universitarios: progreso de la observación a la indagación. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas, 37*(1), 195-211. <https://doi.org/10.5565/rev/ensciencias.2607>

Sampson, V., Enderle, P., y Grooms, J. (2013). Argumentation in science education: Helping students understand the nature of scientific argumentation so they can meet the new science standards. *The Science Teacher*, 80(5), 30.

Sardà, A., y Sanmartí, N. (2000) Enseñar a argumentar científicamente: un reto en las clases de ciencias. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 18(3), 405-422.

Solé Llussà, A., Aguilar, D., Ibañez, M., y Coiduras, J. (2017). Análisis de la comunicación de experiencias indagadoras presentadas en congresos de ciencias dirigidos a alumnos de educación infantil y primaria. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 15(1), 1302(1-13). https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2018.v15.i1.1302

Tamir, P., Nussinovitz, R., y Friedler, Y. (1982). The design and use of a practical tests assessment inventory. *Journal of Biological education*, 16(1), 42-50. <https://doi.org/10.1080/00219266.1982.9654417>

Tytler, R. (2001). Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education*, 23(8), 815-832. <https://doi.org/10.1080/09500690010016058>

Toulmin, S. (2003). *The uses of argument*. Cambridge University Press.

Vila, L., Márquez, C., y Oliveras, B. (2023) Una propuesta para el diseño de actividades que desarrollen el pensamiento crítico en el aula de ciencias. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 20(1), 1302. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2023.v20.i1.1302

Yeh, K., y She, H. (2010). On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context. *Computers & Education*, 55(2), 586-602. <https://doi.org/10.1016/j.compedu.2010.02.020>

Análisis. Categorías de análisis para determinar niveles de desempeño en argumentación.

Nivel	Descripción	Ejemplo
I0 Sin indagación	No hay ninguna intención de investigación, los estudiantes simplemente muestran cómo funciona la App, no hay un diseño experimental o sólo hay una muestra de cómo mide la App.	Los estudiantes solo recogen comentarios de la "App Traductor de perros y gatos", señalan que traduce cosas sin sentidos, pero en ningún momento plantean una investigación.
I1 Indagación deficiente	Los estudiantes plantean un intento de investigación, pero las medidas que obtienen son inválidas y, por lo tanto, los datos y cualquier conclusión que se derive no tienen ningún tipo de validez y fiabilidad.	Los estudiantes usan la "App Medidor de longitudes" sin calibrar y comparar el resultado con una regla de medir, obteniendo una lectura errónea, concluyendo que la App es falsa.
I2 Indagación simple	Los estudiantes plantean un diseño que incluye pocas mediciones, sin ningún elemento adicional como realizar mediciones en el rango completo de la App, usar un número de medidas que permita determinar patrones, considerar la precisión de la App, etc.	Los estudiantes usan la App "Photomath", escanean dos ecuaciones y comprueban que la App las resuelve, pero no exploran la App junto con tomar pocas mediciones.
I3 Indagación Intermedia	Los estudiantes plantean un diseño que incluye varias medidas válidas, agregando un elemento más de investigación: realizar mediciones en el rango completo de la App, usar un número de medidas que permita determinar patrones, considerar la precisión de la App, etc.	Los estudiantes usan la App "Thermal Scanner" y montan 3 experimentos distintos, usan vasos transparentes con agua fría y caliente, vasos de tres colores distintos con agua fría y un radiador de calor. Con ello verifican que la App sólo cambia los colores, sin relación con la temperatura de los cuerpos.
I4 Indagación Compleja	Los estudiantes plantean un diseño que incluye una alta cantidad de mediciones válidas, combinando varios elementos adicionales a la investigación.	Los estudiantes usan la App Sonómetro y controlan el volumen de un televisor mientras registran los decibelios en la App diez veces, determinando un patrón con varios puntos intermedios, luego ponen a prueba la saturación de escala en ambos extremos y de distintas maneras.
A0 Sin argumentación	Los estudiantes no presentan un argumento. Aunque muestren datos, no los utilizan para plantear ninguna argumentación.	Los estudiantes usan la App "Buscador de Fantasmas" solo para recoger comentarios en GooglePlay, y explicar su funcionamiento, pero no usan un argumento sobre la confiabilidad

		de la App.
A1 Argumentación deficiente	La argumentación consta de prueba, razonamiento y asección, pero la prueba presentada es inválida.	Los estudiantes usan la App "Medidor de longitudes" sin calibrar, comparan una medida tomada con la App con la medida realizada con una regla de medir. Utilizan como prueba la comparación entre las medidas. La asección es que la App es falsa, basándose en el razonamiento de que las medidas son distintas.
A2 Argumentación simple	La argumentación consta de prueba, razonamiento y asección, y la prueba es válida y fiable. Sin embargo, no se incluye modulación ni refutación del argumento.	Los estudiantes usan la App "PlantNet", toman dos mediciones comprobando que la App reconoce dos plantas dando por válida la App, sin realizar una modulación o refutación.
A3 Argumentación intermedia	La argumentación, además de prueba, razonamiento y asección, también incluye una modulación sobre la confiabilidad de la App elegida, pero no especifica las restricciones (Refutación) que lo lleva a dicha modulación.	Los estudiantes usan la App "Laser Pointer", escanean veinte veces distintas partes del cuerpo concluyendo que el patrón es aleatorio. La prueba usada es el patrón y la asección es que la App es "rotundamente" falsa.
A4 Argumentación compleja	Además de conectar adecuadamente la prueba, razonamiento y asección, modula su propia conclusión añadiendo las restricciones del contexto.	Los estudiantes usan la App "Detector de Metales", instalan la App en dos teléfonos móviles y comprueban los resultados varias veces. Verifican que la App señala los metales, pero al ser las medidas distintas razonan que tiene restricciones. La asección es que la App es algo verídica, pero que en realidad no es muy fiable.