

Evaluación del pensamiento computacional y sostenibilidad de la propuesta

Acreditar el aprendizaje con criterios explícitos, sin confundir un producto que funciona con un alumno que ha aprendido, y dejar el trabajo montado para el curso siguiente.

- 1 Objeto de la evaluación: producto, proceso y concepto
- 2 Procedimientos de observación de aprendizajes sin registro escrito
- 3 Alcance y limitaciones de las pruebas de pensamiento computacional
- 4 Criterios de calificación y control del efecto del acceso previo a los recursos
- 5 Plan de incorporación de los cambios y calendario de revisión

4.1. Objeto de la evaluación: producto, proceso y concepto

Qué estás evaluando: el producto, el proceso o el concepto.

El robot llega a la meta, el equipo levanta los brazos y tú tienes que anotar algo. Ese es el problema: lo único que has visto es el final. No sabes si llegó porque el grupo razonó la secuencia, porque la fue probando a ojo hasta que salió, porque la escribió el alumno rápido mientras los otros miraban o porque estaba copiada del equipo de al lado. Este epígrafe hace una sola cosa: separar tres objetos que en el aula se confunden a diario y decidir cuál estás mirando. Aquí se decide qué se mira y cuándo; el instrumento con que se recoge es del curso 07.

4.1.1. Tres objetos distintos con el mismo nombre

Cuando en un claustro se dice «evaluar la robótica» se nombran tres cosas a la vez.

El **producto** es lo que queda cuando termina la sesión: el robot en la casilla correcta, el programa que se ejecuta sin pararse, la maqueta acabada. Es observable, es inmediato y es lo que sale en la foto.

El **proceso** es cómo se llegó hasta ahí: cuántas veces se ejecutó la secuencia, qué se cambió después de cada fallo, si el cambio fue razonado o fue mover una tarjeta a ver qué pasaba, y quién tomó las decisiones.

El **concepto** es otra cosa: si el alumno entiende qué es un bucle y lo reconoce en un contexto que no es el de esa actividad —una canción, una receta, el recorrido del autobús—. Se puede tener el producto sin el concepto, y el concepto sin haber terminado nunca el producto.

Los tres son legítimos. Lo que no lo es, que mires uno y anotes otro.

4.1.2. Por qué el producto es el indicador más débil de los tres

El producto que tienes delante funciona por razones que él no declara, y esto no es opinión de aula: lo escriben los autores de las herramientas que lo analizan. Brennan y Resnick (2012) describen el análisis de la cartera de proyectos y declaran su límite: «This approach is entirely product-oriented, and reveals nothing about the process of developing projects» —no revela nada del proceso de desarrollo—, y añaden que mirar los bloques no dice «what the creator was able to do on their own» ni hasta qué punto comprende los conceptos [E]. Quién lo hizo y qué entendió no están dentro del archivo.

Hay una segunda prueba, y viene de los autores de un analizador automático: validaron su puntuación contra métricas clásicas de complejidad del software y se comporta como una medida de complejidad del código, de modo que un proyecto más largo y más lleno de bloques puntúa más, lo haya entendido quien sea. Sus propios autores lo clasifican como instrumento de evaluación **formativa** [E]; calificar con él es usarlo para lo que no sirve. Las cifras de esa validación están en el epígrafe 4.3.

CASO DE AULA · PRIMARIA, TERCER CICLO, MATEMÁTICAS

Tres equipos llevan el robot a la casilla final. El primero escribió la secuencia entera, la ejecutó y corrigió dos giros. El segundo probó ocho veces moviendo una tarjeta cada vez, sin predecir nada. El tercero copió la del primero. El producto es idéntico en los tres: si es lo que entra en la nota, los tres valen lo mismo, y el único que interpretó y modificó un algoritmo —que es el saber básico— es uno.

4.1.3. El proceso: lo que se ve mientras ocurre

El proceso tiene respaldo documentado como objeto de evaluación. Brennan y Resnick (2012) proponen dos vías además del análisis del producto. Las entrevistas basadas en artefactos —conversaciones de sesenta a ciento veinte minutos con treinta y un usuarios de Scratch de ocho a

diecisiete años— permiten una descripción matizada de la comprensión en lugar de un presente o ausente; su límite declarado es que dependen de la memoria del entrevistado y de qué participantes se eligen [E]. Y los escenarios de diseño —el alumno explica qué hace un programa, lo amplía, corrige un error y remezcla una función— los describen los autores como «process-in-action, rather than process-via-memory» [E]. Esa distinción es la que te hace útil el proceso en un aula de veinticinco: no tienes que recordar lo que pasó, tienes que mirar mientras pasa.

Con qué dispositivos concretos se recoge eso dentro de una sesión real, en cuántos momentos y con qué preguntas, es el epígrafe 4.2 entero y aquí no se adelanta.

4.1.4. El concepto: lo que se reconoce fuera de la actividad

El concepto se mide con pruebas, y las pruebas existen: para Secundaria, el *Computational Thinking Test*; para Primaria, dos pruebas en español que no son intercambiables. Cuál usas en cada curso, con sus edades validadas y sus límites, es el epígrafe 4.3 entero. Lo que hace falta aquí es una sola idea, y la declara el propio autor del test: el formato de opción múltiple cerrada mide el concepto en sus niveles bajos, reconocer y entender [E]. Un test de concepto no es un test de proceso, ni pretende serlo, y por eso los tres objetos no se sustituyen unos a otros.

Eso tiene una consecuencia práctica y barata: para comprobar el concepto no hace falta que pases ninguna prueba publicada. Basta con que saques el bucle de su sitio y lo preguntes en otro —la canción que repite el estribillo, el camino de vuelta a casa, la receta que dice «repite tres veces»— y escuches si el alumno lo nombra sin el robot delante. Lo que una prueba te da es una medida comparable con la de otros; lo que tú necesitas casi siempre es saber si el concepto viaja fuera de la actividad en la que se aprendió.

Producto, proceso y concepto: qué deja ver cada uno

	Producto	Proceso	Concepto
Qué es	Lo que queda al acabar: el robot en la meta, el programa que corre	Cómo se llegó: ejecuciones, errores, cambios y decisiones	Si el bucle se reconoce en un contexto distinto
Qué deja ver	Que la secuencia funcionó una vez	Qué falló y qué se hizo con el fallo	Comprensión descontextualizada de la actividad
Qué no dice	Quién lo hizo ni si se entendió	Si el concepto se reconoce fuera de esta actividad	Cómo trabajó el grupo ni quién sostuvo el rotulador
Cuándo se mira	Al final, y solo como indicio	Durante la sesión, mientras ocurre	Antes de empezar y al cerrar
Qué está verificado	El análisis de producto no revela nada del proceso	Proceso en acción, no reconstruido de memoria	Los tests cerrados miden los niveles bajos: reconocer y entender

Fuente: elaboración propia a partir de Brennan y Resnick (2012), Moreno-León, Robles y Román-González (2016) y Román-González (2015). La regla de aula que se deriva es decisión de diseño de este curso, no resultado medido.

FIGURA 4.1

4.1.5. La regla de aula que se deriva, y de qué está hecha

Hasta aquí, lo verificado. Lo que viene es una decisión de diseño de este curso, y se escribe así porque lo es: nadie ha comparado calificaciones de producto y de proceso en programación o robótica escolar [P]. No hay un estudio que diga que evaluar el proceso dé notas más justas. Lo que hay son tres apoyos verificados —el producto no dice quién lo hizo ni si se entendió, la puntuación automática se comporta como medida de complejidad y sus autores la declaran formativa—, y de ahí este curso deriva su regla: **el producto entra como indicio, nunca como prueba principal**.

Indicio quiere decir algo concreto: que el producto confirma lo que ya habías visto durante la sesión y que, cuando lo contradice, manda lo que viste. Si el robot llega a la meta y tú tienes anotado que ese equipo probó ocho veces sin predecir nada, la meta no borra la anotación.

ATENCIÓN · CALIFICAR EL ARCHIVO QUE SE ENTREGA

Poner nota al proyecto terminado es la salida rápida, y la que más se parece a lo que ya hacemos con un cuaderno. Pero el archivo no trae dentro ni la autoría ni la comprensión, y eso lo declaran quienes construyeron las herramientas que lo analizan.

4.1.6. Lo que se decide aquí, y dónde sigue

Decidir es elegir un objeto por momento y escribirlo antes de la sesión. Al principio, concepto: qué sabe el grupo sin haber tocado nada. Durante, proceso: qué error apareció y qué se hizo con él. Al final, concepto otra vez en un contexto distinto, y el producto como indicio. Eso es todo lo que se resuelve aquí. Cómo se observa lo que no deja rastro es el epígrafe 4.2; qué miden y qué no las pruebas publicadas, el 4.3; cómo se califica sin premiar a quien tenía el material en casa, el epígrafe 4.4; y cómo se construye el instrumento, el curso 07.

Aplicación práctica

1. **Levanta la matriz de una actividad tuya.** Toma una actividad con algoritmo, robot o programa dentro que ya hayas puesto en marcha y escríbela en una tabla de tres filas, una por objeto: **lo haces solo**. En cada fila, qué evidencia concreta genera —qué se ve, qué se oye, qué queda escrito—, en qué minuto de la sesión aparece y qué no permite afirmar. Cierra con una línea: cuál de los tres estabas mirando en realidad. Guárdalo como **matriz-producto-proceso-concepto.md**. **Una hora y media.**
2. **Revisa una calificación que ya pusiste.** Elige una nota del curso pasado sobre un trabajo de este tipo y reconstruye de qué está hecha: **lo haces solo**. Anota qué parte venía del producto acabado, qué parte del proceso observado y qué parte del concepto comprobado, y márcalo con lo que sabes hoy: qué entró en esa nota sin que pudieras sostener quién lo hizo ni si se entendió. Termina con la corrección concreta que harías este curso. Guárdalo como **revision-de-una-calificacion.md**. **Una hora y media.**
3. **Escribe tu declaración de qué se mira y cuándo.** Media página lista para pegar en la programación: **lo trabajas con material público**. Copia del currículo de tu comunidad el criterio de evaluación con el que vas a trabajar y debajo escribe los tres momentos —antes, durante y al final—, diciendo en cada uno qué objeto miras y con qué evidencia. Ningún instrumento: ese trabajo es del curso 07. Guárdalo como **que-se-mira-y-cuando.md**. **Una hora.**

4.2. Procedimientos de observación de aprendizajes sin registro escrito



Cómo se observa lo que no deja rastro en el cuaderno.

La sesión ha ido bien. Un equipo se ha atascado en el paso cuatro, lo ha visto, ha tachado una orden y ha vuelto a probar. Otro ha salido a la primera y no sabría explicar por qué. Suena el timbre, se recogen las tarjetas y de todo aquello no queda nada: ni en el cuaderno, ni en la pantalla, ni en tu cabeza el jueves siguiente, cuando haya que poner algo por escrito. Ese problema no se resuelve mirando más, porque mirarlo todo no se puede. Se resuelve decidiendo antes qué vas a mirar, en qué momentos y con qué pregunta.

4.2.1. La observación continua no existe; la programada sí

Con veinticinco alumnos moviéndose, preguntando y tachando a la vez, la observación continua es una promesa que no se cumple ningún día. Hay que hacer lo contrario de ampliar la mirada: recortarla. Tres o cuatro momentos, escritos antes de entrar, con un objeto concreto cada uno.

Un dato respalda ese recorte. El metaanálisis de Südkamp, Kaiser y Möller (2012) reunió setenta y cinco estudios y encontró una correlación media de $r = 0,63$ entre el juicio del docente y el rendimiento medido con prueba estandarizada [E]: razonablemente bueno y claramente imperfecto, comparte alrededor del 40 % de la varianza con la medida objetiva. Y el moderador que importa da la regla de trabajo: el juicio **informado** —cuando quien observa sabe de antemano qué se le pregunta y sobre qué— se corresponde mejor con el rendimiento real. No es un estudio sobre robótica ni sobre programación escolar; lo que autoriza es saber qué buscas antes de buscarlo.

4.2.2. Los cuatro dispositivos que caben en una sesión real

La referencia de fondo son los tres enfoques de Brennan y Resnick que viste en el epígrafe 4.1, cada uno con su límite declarado: la cartera de proyectos no dice nada del proceso, la entrevista depende de la memoria del entrevistado y el escenario de diseño es lo único que ocurre delante de ti mientras pasa [R]. De ahí salen los cuatro dispositivos, y el criterio para elegirlos es ese: lo que se puede mirar dentro de la sesión, no lo que se reconstruye después.

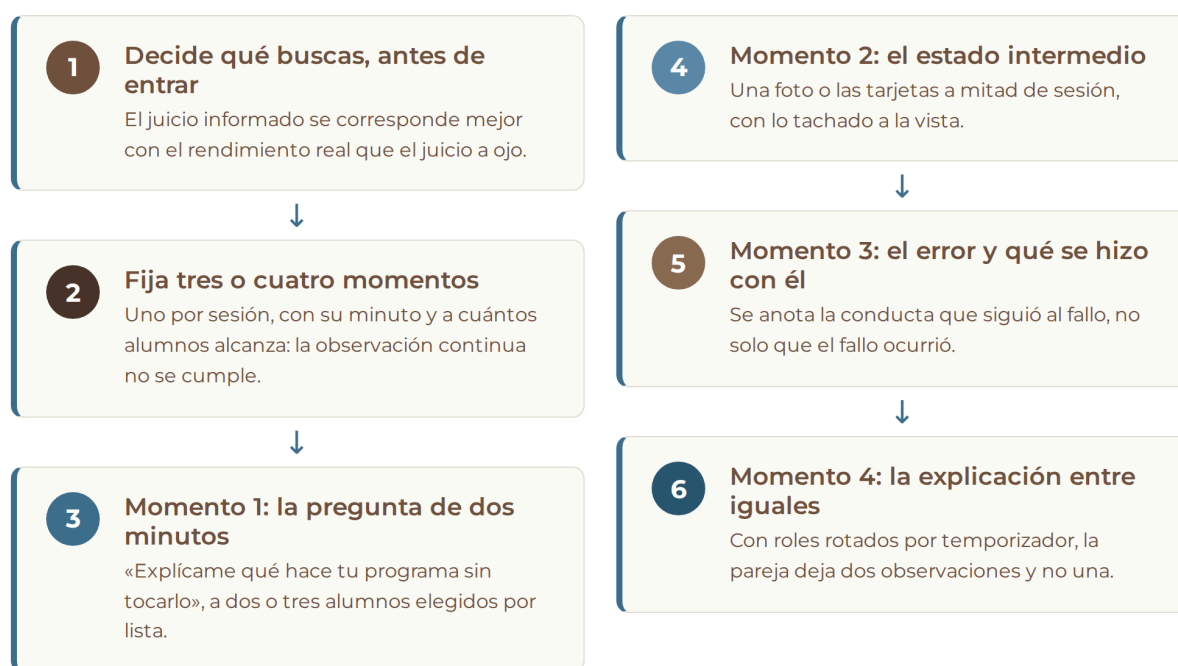
La pregunta de dos minutos. «Explícame qué hace tu programa sin tocarlo.» Sin tocarlo es la parte que trabaja: con las manos quietas no cabe sustituir la explicación por una demostración. Se hace a dos o tres alumnos por sesión y se anota en una línea si describe el propósito, si describe los pasos o si solo señala.

La captura del estado intermedio. Una foto o las tarjetas de órdenes tal como están a mitad de sesión, con lo tachado a la vista. El producto final borra la historia; el estado intermedio la conserva.

El registro del error y de qué se hizo con él. No basta con anotar que falló: lo evaluable es la conducta que siguió. Volvió a leer, probó al azar, cambió una orden y comprobó, o llamó.

La explicación entre iguales. Un alumno le cuenta a otro cómo resolvió y tú escuchas sin intervenir. Es el más barato de los cuatro y el que más rápido separa a quien entendió de quien copió el bloque del equipo de al lado.

Una observación programada, paso a paso



Fuente: elaboración propia a partir de Brennan y Resnick (2012), Südkamp y otros (2012), Sun, Yang y Becker (2024) y Childs y Sentance (2024).

FIGURA 4.2

4.2.3. La depuración se mira porque se enseña

Tratarla como objeto de evaluación y no como contratiempo tiene respaldo, y es el del epígrafe 2.1: el metaanálisis de Sun, Yang y Becker (2024) sitúa el efecto medio de enseñar a depurar en $g = 0,64$, con un intervalo que va de $0,32$ a $0,96$ [E]. Lo que importa aquí son las dos puntas de ese intervalo: la instrucción sistemática del procedimiento funciona y mejorar los mensajes de error no sirve de nada. Sirve para esperar algo, no para prometerlo.

Si la depuración se enseña y mejora, es un aprendizaje, y un aprendizaje se mira: el minuto en que un equipo descubre el fallo deja de ser el minuto perdido y pasa a ser el momento que tenías programado observar.

4.2.4. Qué aprendió el equipo y qué aprendió cada alumno

En trabajo por parejas y por equipos, este es el punto ciego: lo que ves es conjunto y lo que tienes que acreditar es individual, y entre una cosa y la otra no hay paso automático.

Hay una práctica documentada que ayuda, y se cuenta entera. Childs y Sentance (2024) estudiaron la programación por parejas en Primaria y su resultado principal fue nulo: no encontraron efecto estadísticamente significativo sobre la actitud de las alumnas hacia la informática — $0,046$ puntos en una escala de uno a cuatro, $p = 0,331$ — ni sobre la intención de seguir estudiándola — $4,3$ puntos porcentuales, $p = 0,453$ — [E]. Se probó y no se encontró. Lo aprovechable es otra cosa: la rotación de roles funcionó cuando se gestionó de forma explícita, con cambio cada cinco minutos marcado por un temporizador. Eso convierte la pareja en dos observaciones y no en una. Cada alumno pasa por los dos papeles delante de ti, y la pregunta de dos minutos se le hace al que no tenía las manos en el material.

ATENCIÓN · EL ALUMNO QUE EXPLICA MEJOR NO ES EL QUE MÁS SABE

La observación premia sin querer a quien verbaliza con soltura, y en un aula donde casi todo lo evaluable se dice en voz alta ese sesgo tiene vía libre. Es recomendación profesional de este curso [P] y no un hallazgo: no se ha encontrado evidencia específica sobre ello en programación ni en robótica escolar. La contramedida es de oficio: elige antes a quién preguntar —por lista, no por quien levante la mano— y acepta la explicación que se da señalando el material.

4.2.5. Qué se decide aquí y qué no

Este epígrafe decide qué se busca y en qué momento. Cómo se convierte eso en rúbrica, lista de cotejo o escala es del curso 07, y allí se remite; la organización del proyecto largo, del curso 08. Y queda una decisión práctica: tres o cuatro anotaciones por sesión y por equipo, escritas en el momento y no reconstruidas por la tarde, sostienen una valoración de proceso al final de la secuencia. Lo que no se anotó no pasó.

CASO DE AULA · EDUCACIÓN INFANTIL, SEGUNDO CICLO

Tarjetas de flecha sobre la cuadrícula del suelo, cuatro sesiones. Momentos programados: en la primera, quién coloca las tarjetas y quién mira; en la segunda, la pregunta de dos minutos —«cuéntame qué va a hacer el muñeco»— a tres niños elegidos por lista; en la tercera, la casilla de más y qué hizo cada equipo al ver que se pasaba; en la cuarta, la explicación entre iguales. Lo que se lleva la docente: una foto de la cuadrícula a mitad de la tercera sesión y once líneas de anotación. De los dos equipos que acabaron bien, uno no supo explicar el paso que había cambiado.

Aplicación práctica

1. **Construye tu plan de observación programada.** Toma una secuencia tuya de tres a cinco sesiones —vale la del 3.1— y fija tres o cuatro momentos de observación, uno por sesión: **lo haces solo**. De cada momento escribe qué buscas, en qué minuto cae, a cuántos alumnos alcanza y con qué frase lo provocas. Añade la hoja de registro con la que vas a entrar en clase: una cara de folio, momentos en filas y equipos en columnas, para escribir allí y no de memoria; convertirla en rúbrica o en escala es del curso **07**. Guárdalo como [plan-de-observacion.md](#). **Dos horas.**
2. **Aplica el plan en una sesión real y registra lo que pasó.** Lleva el plan a una sesión con tu grupo y úsalo: **lo haces solo**. Si el centro tiene robots o tabletas disponibles, obsérvate una sesión con material; si no los tiene, o están sin pilas, la sesión que se observa es la secuencia desenchufada del 3.1, con el mismo plan y sin cambiar nada —alumno-robot con tarjetas, receta ejecutada literalmente o dibujo por coordenadas—, porque los cuatro dispositivos no dependen del aparato: la pregunta de dos minutos se hace igual sobre una tarjeta de órdenes que sobre una pantalla, y el estado intermedio se fotografía igual. Al terminar, escribe lo observado equipo por equipo y, aparte, lo que falló del plan: el momento que no llegaste a mirar, la pregunta que no dio nada y el minuto que se comió la recogida del material. Guárdalo como [sesion-observada.md](#). **Dos horas.**
3. **Escribe los dos guiones que vas a usar cada vez.** Redacta el de la pregunta de dos minutos: **lo haces solo**. La frase de apertura, las dos repreguntas para cuando la respuesta se queda en señalar la pantalla y las tres respuestas tipo que vas a distinguir —describe el propósito, describe los pasos, solo señala—. Debajo, el de la explicación entre iguales: cómo formas los pares, cuánto dura, qué le pides al que escucha y qué haces tú mientras. Guárdalo como [guiones-de-pregunta-y-explicacion.md](#). **Una hora.**

4.3. Alcance y limitaciones de las pruebas de pensamiento computacional



Qué miden los tests de pensamiento computacional y qué no.

Alguien del equipo de digitalización ha encontrado un test de pensamiento computacional en internet, lo ha traducido y te propone pasarlo en tercero para tener un dato antes de la evaluación. Es razonable, y puede salir mal por tres motivos: que el instrumento no esté validado para esa edad, que lo esté y aun así no discrimine ahí, o que mida algo que no querías mirar. Antes de contestarle, mira qué prueba es, para qué edad se validó y qué declara no medir. Eso ordena la decisión, y también dice lo que ninguno de los instrumentos acredita hoy.

4.3.1. El CTt: qué mide, para qué edad y qué declara su propio autor

Si das clase en Secundaria, la prueba de referencia en español es el *Computational Thinking Test* (CTt) de Román-González (2015): 28 ítems de opción múltiple, 45 minutos, aplicación colectiva, sobre siete conceptos, de las secuencias a los tres tipos de condicional y las funciones simples [E]. Su población de diseño son los 12 y 13 años, 1.º y 2.º de ESO; el estudio de validez lo aplicó de 5.º de Primaria a 4.º de ESO, y ese es el rango con respaldo: fuera de ahí no lo tienes validado. Lo más útil es lo que declara el autor: el formato de opción múltiple cerrada hace que el test «only measures CT at their lower levels of complexity (“Recognize” and “Understand”)», y él mismo se pregunta por escrito si no será «“just” an AT-test (Algorithm Thinking Test)». Medir los niveles superiores exige problemas abiertos, y eso un test colectivo no lo hace.

4.3.2. Primaria: dos pruebas en español, y no son intercambiables

Si das Primaria tienes dos, y la diferencia entre ellas decide tu octubre. El **BCTt** (Zapata-Cáceres, Martín-Barroso y Román-González, 2020) son 25 ítems de opción múltiple, desenchufados, para 1.º-6.º, con $\alpha = 0,82$ sobre 200 alumnos de validación [E]. El **cCTt** (El-Hamamsy y otros, 2022) son también 25 ítems desenchufados, pero para 3.º y 4.º (7-9 años), con $\alpha = 0,85$ sobre 1.519 alumnos de 77 aulas del cantón suizo de Vaud [E]; mide conceptos, no prácticas ni perspectivas.

Entre las dos decide un hallazgo publicado. El-Hamamsy y otros (2022), en *Frontiers in Psychology*, las compararon sobre 575 alumnos de 3.º y 4.º de Portugal y encontraron que **el BCTt presenta efecto techo en esos cursos**: 13 de sus preguntas superan el umbral máximo del índice de dificultad, y solo discrimina en el tramo de baja capacidad, mientras que el cCTt no lo presenta. La recomendación es literal: «For these grades, the BCTt can thus be employed as a screening mechanism to identify low ability students» [E].

ATENCIÓN · UN TECHO SE DECLARA, NO SE HEREDA

Ese efecto techo es un resultado publicado sobre el BCTt en 3.º y 4.º de Primaria. No te autoriza a decir que «los tests de pensamiento computacional se saturan», ni a suponerlo del cCTt, del CTt o de una prueba traducida. Cada instrumento declara sus edades, su muestra y sus límites, y los copias antes de pasarlo.

4.3.3. Los primeros cursos: una prueba desenchufada y un alfa que hay que mirar

Si tu grupo está entre los 5 y los 9 años, lo que hay es TechCheck (Relkin, de Ruiter y Bers, 2020): prueba desenchufada de opción múltiple, que pasas al grupo entero, sobre seis dominios, de los algoritmos a la depuración. Son 15 ítems puntuables más dos de práctica, con una duración media real de 13,40 minutos, sobre 768 alumnos de 1.º y 2.º de Primaria [E]. Su α de Cronbach es 0,68: bajo, así que no la trates como medida de precisión individual. Y en su publicación original no estaba validada en Infantil. Sus autores declaran además posible efecto techo —el 8 % sacó la

puntuación máxima— y que no hay patrón de oro a esas edades. Para Infantil hay una adaptación posterior de la misma prueba, sin cifras verificadas aquí: si la usas, copia sus edades y su muestra antes de pasarla.

4.3.4. Bebras y el análisis automático: dos usos legítimos que no son calificar

El reto Bebras nació en 2004 en Lituania, de la mano de Valentina Dagienė: 15 tareas en 40 minutos, de opción múltiple y sin conocimientos previos de informática. Su único dato de alcance verificado es de 2014 —33 países y más de 900.000 participantes—, y el uso que el profesorado le da, según sus promotores, es orientar y detectar, no poner nota [E]. Y hay cautela medida: un análisis con teoría de respuesta al ítem (Oliveira y otros, 2025) encuentra que la dificultad asignada a priori a cada tarea acierta entre el 53 % y el 58 % de las veces, y que las tareas se concentran de fácil a medio: discrimina mal arriba.

Dr. Scratch (Moreno-León, Robles y Román-González, 2015) puntúa siete dimensiones de 0 a 21 y está declarado por sus autores como herramienta **formativa**. Esos mismos autores lo validaron después (2016) contra métricas de complejidad del software sobre 95 proyectos, con correlaciones significativas con la complejidad ciclomática de McCabe ($r^2 = 0,385$) y con el vocabulario (0,461) y la longitud (0,319) de Halstead [E]: se comporta como una medida de complejidad del código. Es el límite del análisis de producto que viste en el epígrafe 4.1: un proyecto remezclado puntúa igual que uno propio, y calificar con él es usarlo para lo que no sirve [P].

Los tres se complementan, y está medido cuánto: $r = 0,52$ entre el CTt y un conjunto de tareas Bebras ($n = 179$), y $r = 0,44$ predictivo y $r = 0,53$ concurrente entre el CTt y Dr. Scratch ($n = 71$) (Román-González, Moreno-León y Robles, 2017) [E]. Si pasas dos, no estás midiendo dos veces lo mismo: tres cuartas partes de lo que ve cada uno no lo ve el otro.

4.3.5. Qué parte de lo que miden es específica de esta materia

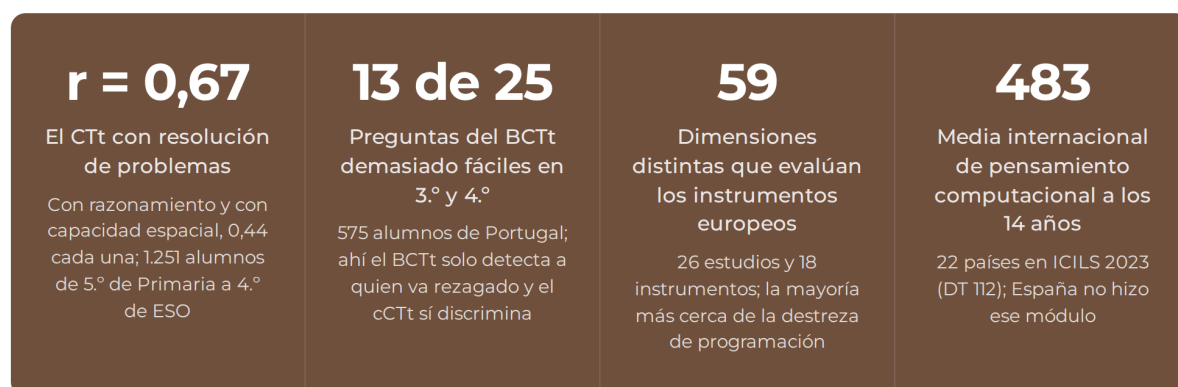
En Román-González, Pérez-González y Jiménez-Fernández (2017), sobre 1.251 estudiantes de 5.º de Primaria a 4.º de ESO, la puntuación del CTt correlaciona $r = 0,67$ con resolución de problemas —fuerte—, pero $r = 0,44$ con razonamiento y $r = 0,44$ con capacidad espacial, ambas moderadas [E]. Los autores concluyen que eso respalda tratarlo como capacidad de resolución de problemas. Queda abierto cuánto de lo que mide es específico de esta materia, y por eso lees esto junto al epígrafe 1.3.

El estado del campo cierra el argumento: la revisión de Babazadeh y Negrini (2022) examina 26 estudios europeos con 18 instrumentos y encuentra 59 dimensiones distintas; de ellas, «the vast majority [...] were closer to programming skills rather than CT per se». Y declara dos huecos que te afectan: no hay definición operativa común, y **no existe un modelo de competencia que diga qué debe alcanzar un alumno a cada edad** [R].

4.3.6. El estado de los datos: la medida internacional existe y España no está en ella

España participó por primera vez en ICILS en 2023, con 11.800 estudiantes de 2.º de ESO de 508 centros, y obtuvo 495 puntos en la escala CIL —que no es la competencia digital del perfil de salida—, frente a 493 de la Unión Europea y 476 de media de los participantes [E]. Pero no hizo el módulo opcional de pensamiento computacional. Sí lo hicieron 22 países, con una media equiparada de 483 puntos (DT 112) y un recorrido de 127 puntos entre el primero y el último. Así que no digas «no hay datos»: la medida existe a los catorce años y España no está en ella. La primera será ICILS 2028, con el alumnado que hoy tienes en 6.º.

Cuatro cifras sobre lo que miden los tests



Fuente: elaboración propia a partir de Román-González, Pérez-González y Jiménez-Fernández (2017), El-Hamamsy y otros (2022), Babazadeh y Negrini (2022) e IEA, ICILS 2023.

FIGURA 4.3

CASO DE AULA · 2.º DE ESO, TECNOLOGÍA Y DIGITALIZACIÓN

Un dato de partida en octubre y otro en mayo, el CTt en los dos: 2.º está dentro de su población de diseño, y una prueba traducida sin edades declaradas no. Lo que acredita es reconocimiento de conceptos, no lo que cada alumno sabe hacer solo; el instrumento con el que se califica es del curso 07.

Aplicación práctica

1. **Elige el instrumento para tu curso y justifica la elección por escrito.** Sobre el curso que tienes este año, escribe una página con la prueba elegida, su publicación, sus edades validadas, sus ítems y su duración: **lo haces solo**. Incluye un párrafo sobre el efecto techo del BCTt en 3.º y 4.º y su consecuencia en tu decisión, y otro con qué harías si tu curso queda fuera de los rangos validados. Guárdalo como [eleccion-de-instrumento.md](#). **Una hora y media.**
2. **Lee diez tareas tipo Bebras y clasifica lo que mide cada una.** El banco de tareas está accesible todo el año, no solo durante la convocatoria: **lo trabajas con material público**. Elige diez del grupo de edad que te corresponda y anota, para cada una, qué operación pone en juego —descomponer, reconocer patrones, abstraer, algoritmizar o depurar— y en qué nivel de dificultad la colocarías tú. Cierra diciendo cuántas etiquetas oficiales de dificultad cambiarías. Guárdalo como [bebras-diez-tareas.md](#). **Una hora y media.**
3. **Escribe qué no acredita la prueba que has elegido.** Media página, guardada junto al instrumento: **lo haces solo**. Di qué queda fuera —el proceso, la autoría, las prácticas y perspectivas, los niveles altos de complejidad— y qué decisión de aula **no** vas a tomar con ese número delante. Guárdalo como [que-no-acredita-la-prueba.md](#). **Una hora.**

4.4. Criterios de calificación y control del efecto del acceso previo a los recursos



Cómo se califica sin premiar al que tenía el kit en casa.

Dos coches teledirigidos llegan a la exposición de final de trimestre. Uno lleva chasis impreso en tres dimensiones, luces y carcasa pintada; el otro es cartón, gomas y un motor recuperado. Los dos completan el circuito. Si lo que puntúa es el coche, ya sabes cuál gana, y no habrás medido pensamiento computacional: habrás medido qué había en el salón de cada casa. Este epígrafe va de eso: de dónde se cuela la desigualdad dentro de una nota y qué decisiones la sacan. Empieza corrigiendo dos tópicos, porque el argumento de equidad se apoya casi siempre en cifras que aquí no dicen lo que se cree.

4.4.1. Dos tópicos que hay que quitar de en medio antes de argumentar

El primero es que la brecha española sea especialmente grave. No lo es. En ICILS 2023, la diferencia a favor del alumnado con mayor índice socioeconómico y cultural fue en España de 69 puntos de escala, frente a los 79 del promedio de la Unión Europea [E]. La brecha existe y es grande, y está por debajo de la europea; escribirlo al revés falsea el dato, y un dato falseado se cae en la primera reunión en que alguien lo mire.

El segundo tópico es el que más se oye: «hay niños sin internet en casa». En 2026 esa frase no se sostiene en España. Según la encuesta del Instituto Nacional de Estadística difundida el 20 de noviembre de 2025, el 97,4 % de los hogares disponía de acceso a internet [E]. Si montas tu política de aula sobre la conexión, el primero que abra esa nota te deja sin argumento, y con él se va también lo que sí era cierto.

Lo verificable está en otro sitio, y lo señalan dos cifras del mismo ICILS: tener conexión en casa valía 22 puntos en España (26 en la Unión Europea), pero disponer de dos o más ordenadores valía 32 (39 en la Unión Europea) [E]. No es el acceso de la vivienda: es el aparato que el niño puede usar cuando quiere, sin esperar turno. Y con él, el material del kit, el tiempo y la ayuda adulta de las tardes.

4.4.2. Lo que viene ahora es diseño, y se dice antes

Va en la primera línea: las decisiones de calificación que siguen las toma este curso, y no hay ningún resultado detrás que las obligue [P]. Nadie ha comparado calificaciones de producto y de proceso en robótica o programación escolar segmentando por nivel socioeconómico. Lo que hay son tres apoyos por analogía, y como analogía se citan.

El primero, Rønning (2011), sobre alumnado de Primaria de los Países Bajos: la diferencia de resultados es mayor en las clases donde se mandan deberes a todos, porque el alumnado desfavorecido recibe menos ayuda de sus familias con esos deberes [E]. El segundo, la OCDE (2014), con datos de PISA: el alumnado de entorno favorecido dedica más tiempo al trabajo fuera del aula [E]. Los dos son de deberes, no de robótica. El tercero es el precedente más directo que existe: en 2019, el regulador inglés Ofqual retiró de la nota el proyecto de programación del título de Secundaria de Informática porque no podía garantizar de quién era el trabajo, y aun así lo mantuvo obligatorio, veinte horas lectivas, sin calificación [E]. Con su letra pequeña: el motivo declarado fue la filtración de soluciones a foros de internet, no la desigualdad de recursos. Sostiene que calificar un producto hecho sin supervisión no es defendible; no vale como medida de equidad.

4.4.3. Las cuatro decisiones que sacan la casa de la nota

El trabajo que puntúa se termina dentro del aula. Puede haber trabajo fuera, y puede ser estupendo; lo que no puede es entrar en la nota, porque fuera no controlas el tiempo, la ayuda ni el material. Es la arquitectura de Ofqual: se hace en horario y la nota no cuelga de eso.

Se valora el avance sobre el punto de partida. Quien llega sabiendo poner un bucle porque lleva dos años con un kit en casa parte de un sitio; quien lo descubre en tu clase, de otro. Midiendo solo el punto de llegada, mides la casa.

El componente comprado no puntúa. Ni el chasis impreso, ni las luces, ni la carcasa. Se declara por escrito antes de empezar, porque dicho después parece un castigo.

Acabado y solución se separan. El acabado puede reconocerse, y en un reto STEAM tiene su sitio, pero va en su propia línea y no arrastra la nota del algoritmo: es la parte que más depende del bolsillo.

Qué parte de la nota depende de lo que hay en casa

Calificar sin premiar el kit de casa

Lo que dicen los datos

- Brecha por nivel socioeconómico: 69 puntos en España
- Promedio de la Unión Europea: 79 puntos
- Dos o más ordenadores en casa: 32 puntos
- Hogares con internet en España: 97,4 % (INE)

Lo que se reparte mal

- El aparato disponible para el niño
- El material del kit
- El tiempo fuera del aula
- La ayuda adulta por las tardes

Cuatro decisiones de diseño

- Lo que puntúa se termina en el aula
- Se valora el avance sobre el punto de partida
- El componente comprado no puntúa
- Acabado y solución van en líneas distintas

Lo que sostiene cada cosa

- Ofqual retiró de la nota el proyecto, no la práctica
- Rønning y la OCDE son de deberes, no de robótica
- Una parte individual dentro de la nota de grupo
- El rendimiento por sexo no es el problema: curso 16

Fuente: elaboración propia a partir de ICILS 2023 (INEE, 2024), INE (2025), Rønning (2011), OCDE (2014), Ofqual (2019), Childs y Sentance (2024) y Comisión Europea (2019).

FIGURA 4.4

ATENCIÓN · LA EXPOSICIÓN DE FINAL DE TRIMESTRE

La desigualdad entra sin que nadie la invite por la muestra pública del producto acabado: nadie la puntúa oficialmente y todo el mundo la puntúa. Si vas a hacerla —hazla, vale la pena—, pide a cada equipo que explique delante de su trabajo una decisión de programación y qué falló y cómo lo arreglaron. Así lo que se expone es el proceso, y el cartón y la impresora tres D empiezan en el mismo sitio.

4.4.4. La nota de grupo y la parte que tiene que ser individual

El trabajo en equipo mezcla lo que aprendió cada uno con lo que aprendió el grupo, y aquí mezcla además lo que traía cada uno de casa. La regla de este curso también es diseño, no hallazgo: una parte de la calificación tiene que apoyarse en algo que el alumno produce por su cuenta [P]. Basta con poco: explicar sin tocar qué hace el programa, señalar en qué línea está el fallo, rehacer solo un fragmento de lo que el equipo resolvió junto.

Ayuda la única práctica documentada que hay aquí, la rotación temporizada de roles que viste en el epígrafe 3.3 y en el 4.2: cada alumno pasa por los dos papeles delante de ti. Reparte el teclado; el mismo estudio comprobó que no cura nada más.

El registro con que se recoge todo esto —rúbrica, lista de cotejo, escala, peso de cada parte— es del curso 07, y allí se remite. Y si lo que tienes delante no es una diferencia de material sino una barrera de acceso, eso es del curso 06.

4.4.5. El estereotipo que más daño hace, con el dato europeo delante

Queda una desigualdad que entra por otro sitio, y sobre ella hay dato bueno. No es cierto que los chicos rindan más en pensamiento computacional. El informe de la Comisión Europea sobre ICILS 2018 lo dice con la letra: en competencia informacional y digital las chicas puntuaron significativamente más alto en todos los Estados miembros participantes, de 11 puntos en Chequia y Portugal a 29 en Eslovenia y Finlandia; en pensamiento computacional las diferencias fueron distintas, pequeñas e inconsistentes, y solo significativas en Finlandia —chicas, 13 puntos— y Portugal —chicos, 16 puntos— [E]. Signos opuestos: no es un patrón de capacidad.

Lo sistemático y grande no es el rendimiento: hay evidencia estadounidense de que el estereotipo sobre el interés de las niñas por la informática está instalado muy pronto y se asocia a menos interés declarado [E]. Cómo se trabaja eso no es de este curso. Interés, autoeficacia y elección: ahí está la diferencia real, y su tratamiento en la orientación hacia lo técnico es del curso 16.

Aplicación práctica

1. **Revisa una calificación tuya en clave de equidad.** Coge la última nota que hayas puesto a un trabajo con producto —robótica, programación o reto STEAM— y desglósala en las partes que de verdad sumaron: **lo haces solo**. Marca en cada parte si dependía del tiempo, del material o de la ayuda disponibles fuera del centro, y calcula qué porcentaje de la nota dependía de eso. Guárdalo como [revision-de-nota-en-clave-de-equidad.md](#). **Una hora y media.**
2. **Diseña el trabajo terminado dentro del aula y la parte individual de la nota de grupo.** Reescribe esa propuesta para que todo lo que puntúa ocurra en horario lectivo: **lo haces solo**. Fija cuántas sesiones hacen falta, qué pasa si un equipo no termina, qué componente comprado queda fuera de la nota y en qué línea aparte va el acabado. Define después la parte individual: qué produce cada alumno por su cuenta y en qué minuto de qué sesión se observa. El peso que esa parte tenga dentro de la calificación no se decide aquí: es del **curso 07**. Guárdalo como [trabajo-en-aula-y-parte-individual.md](#). **Dos horas.**
3. **Escribe tu respuesta a las dos preguntas de equidad más frecuentes.** Media página por pregunta, con las cifras delante: **lo trabajas con material público**. La primera, la de la familia que dice que su hijo está en desventaja porque en casa no tiene material: contesta con los 69 puntos de España frente a los 79 de la Unión Europea en ICILS 2023, con el 97,4 % de hogares con internet del INE y con lo que sí se reparte mal, y cierra con lo que has decidido hacer en la nota. La segunda, la de quien afirma que los chicos se apañan mejor con esto: contesta con el dato europeo de ICILS 2018 y con lo que ese dato no dice. Guárdalo como [dos-respuestas-de-equidad.md](#). **Una hora.**

4.5. Plan de incorporación de los cambios y calendario de revisión

Qué te llevas: tres decisiones, su coste y su punto de revisión.

Al terminar un curso como este hay dos salidas. Una es cerrar el manual con veinte ideas apuntadas y ninguna escrita en la programación, que es lo que ocurre casi siempre, y no por falta de ganas: veinte ideas no caben en septiembre. La otra es salir con tres decisiones, solo tres, cada una con lo que cuesta prepararla, el escalón de equipamiento que exige y un indicador que puedas mirar en tu aula en una fecha concreta. Este epígrafe monta esa salida, deja escritos los guiones de las dos conversaciones que siempre llegan y fija qué se hereda.

4.5.1. Por qué tres, y qué tiene que llevar dentro cada una

Tres es el número que cabe en un curso escolar sin robarle sitio a lo que ya das. Ningún estudio fija ese número; lo fija la experiencia de quien escribe esto [P]. Tres decisiones preparadas rinden más que diez enunciadas, porque cada una de las diez volvería a improvisarse en el aula.

Una decisión escrita lleva cuatro cosas, y si le falta alguna no es una decisión, es una intención. El enunciado: qué vas a hacer distinto, en qué grupo y sobre cuál de los saberes básicos del epígrafe 1.2. El coste en horas de preparación, contado sobre lo que cuesta producir el material, no sobre lo que dura la sesión. El escalón de equipamiento, con el papel de partida y el material solo donde aporte lo que el papel no da. Y el indicador observable: qué vas a mirar, cuándo y con qué registro, que es lo del epígrafe 4.1 y el 4.2.

4.5.2. Los puntos de revisión, y que sean fechas y no propósitos

Un indicador sin fecha no se mira nunca. Por eso cada decisión se ata a un punto del calendario real, y esos puntos no son momentos de reflexión: son preguntas cerradas. En octubre, si la secuencia entró donde dijiste; en diciembre, qué falló; en marzo, si subió el escalón; en junio, qué queda escrito.

Los cuatro puntos de revisión del curso que viene

Junio	●	Se eligen tres Tres decisiones y no más, cada una con su coste en horas y su escalón.
Septiembre	●	Entran en la programación Cada decisión queda atada a un saber básico y a un grupo concreto.
Octubre	●	¿Entró donde dijiste? Se comprueba si la secuencia ocupó la semana prevista o se desplazó.
Diciembre	●	Qué falló, por escrito Único punto donde una decisión se retira; lo que no entró no se arrastra.
Marzo	●	¿Subió el escalón? Se anota si el material apareció y qué se hizo con la versión de papel.
Junio	●	Qué se hereda Documento de continuidad depositado, con nombre de quien lo recoge.

Fuente: elaboración propia.

FIGURA 4.5

La revisión de diciembre es la única que puede cambiar una decisión a mitad de camino: lo que a esas alturas no ha entrado en el aula no se arrastra hasta junio, se retira y se anota por qué.

4.5.3. La conversación con dirección: qué ha dado de sí el dinero de los robots

Suele llegar en un claustro de junio. La respuesta honrada no es que el dinero se perdió ni que valió la pena: es que **la pregunta, así formulada, no tiene respuesta con lo que hay**. Se acredita lo que se hizo, con quién, cuántas sesiones y qué se observó; no se acredita un retorno, porque nadie lo ha medido en tu centro.

Por qué no se sabe en general —el ritmo al que cambian los productos, quién financia la evidencia y lo poco que la exige quien compra— es materia del curso 04, que trae las cifras y el informe internacional del que salen. Aquí basta con saber que existen y con no inventar ninguna. Lo que sí es de este curso es el par de cifras que el metaanálisis de Scherer, Siddiq y Sánchez Viveros (2019) dejó medido, y ordena la reunión mejor que cualquier otra cosa: frente a no hacer nada, programar da $g = 0,65$; frente a otra actividad equivalente, $g = 0,16$ [E]. Con el robot o sin él, lo que se compara no es «tecnología sí o no», sino esta actividad contra la que se daría en su lugar. Esa es la pregunta que tu centro sí puede contestar, y la contesta con el registro de lo que hicisteis.

Esa prudencia tiene un antecedente con nombre. Cuban, en *Oversold and Underused* (Harvard University Press), describió centros donde los equipos se usaban mucho menos en el aula que en casa y, cuando se usaban, poco y de forma no creativa. Es trabajo de campo estadounidense de finales de los noventa, anterior a la tableta y a la banda ancha: se cita como antecedente y nunca como evidencia de hoy. De él sigue en pie su condición de partida: la implantación funciona cuando el profesorado domina la herramienta, cree que mejora el aprendizaje y conserva autonomía curricular.

ATENCIÓN · CONTESTAR CON UNA CIFRA DE RETORNO

La tentación es dar un número —«ha subido el interés», «han mejorado en matemáticas»—, porque es lo que la pregunta pide. Ese número no existe y quien lo dé lo tendrá que sostener el año siguiente. Se entrega un documento: qué grupos, qué sesiones, qué se observó, qué falló y qué se hará distinto. Se acredita el trabajo, no el retorno.

4.5.4. La conversación con la familia: «he oído que programar mejora las Matemáticas»

Aquí las cifras son las del epígrafe 1.3 y se dan enteras. El mismo metaanálisis, con 105 estudios, encuentra transferencia — $g = 0,49$ — repartida de forma muy concreta: la cercana por encima de la lejana ($g = 0,75$ frente a $g = 0,47$), con los beneficios en pensamiento creativo, destrezas matemáticas, metacognición, razonamiento y habilidades espaciales, con el rendimiento escolar beneficiándose poco ($g = 0,28$) y con la competencia lectora en un efecto indistinguible de cero [E]. Y delante de todas, el par que acabas de usar en el claustro: $0,65$ frente a no hacer nada, $0,16$ frente a otra actividad equivalente.

Traducido a una tutoría: hay algo, es pequeño, va sobre todo a lo que se le parece, y la mayor parte de lo que se atribuye a programar aparece porque el grupo de comparación no hacía nada. No se promete mejora en matemáticas ni en lectura, ni que enseñe a pensar en general: se explica qué va a hacer el niño, qué se va a observar y cuándo. Y si la pregunta va al futuro laboral, el porcentaje de empleos que aún no existen que circula por las presentaciones del ramo no tiene fuente identificable.

«¿Le va a servir para Matemáticas?» La respuesta cabe en tres frases: se ha medido transferencia a destrezas matemáticas, es pequeña y crece cuando la actividad se parece a lo que luego se le pide; aquí vamos a interpretar y modificar secuencias, que es lo que su curso pide; y voy a mirar cómo depura cuando algo no sale, no si el resultado funciona.

4.5.5. El plan de continuidad: qué queda escrito y quién lo hereda

Lo que no está escrito no se hereda. El documento de continuidad son dos páginas con cinco entradas. Las tres decisiones, con su resultado de junio. Los materiales producidos y dónde están: las tarjetas del epígrafe 3.1, el protocolo del 3.3, la hoja de registro del 4.2. Qué falló, con fecha. El acuerdo de coordinación del epígrafe 3.5. Y quién lo recoge, con el sitio donde queda depositado, que es la programación de equipo o de departamento y no una carpeta personal.

Hay una fecha que ayuda a defender esa continuidad ante un equipo directivo, y está en el epígrafe 4.3: la primera medición internacional española de pensamiento computacional será ICILS 2028, y el alumnado que se evaluará entonces está hoy en 6.º de Primaria [R]. No es un argumento de resultados, es de calendario.

4.5.6. Lo que no se decide aquí

El formato del proyecto largo —pregunta guía, producto público, equipos, calendario— es del curso 08; los instrumentos, del curso 07; la situación de aprendizaje, del 05; las plataformas, las cuentas y la protección de datos, del 04; las barreras de acceso, del 06; los puntos, las insignias y las narrativas, del 09; las altas capacidades, del 13; la orientación hacia lo técnico y su sesgo por sexos, del 16; toda explicación apoyada en hemisferios o pirámides de retención, del 17; y la Formación Profesional, del 20.

Aplicación práctica

1. **Escribe tus tres decisiones para el curso que viene.** Una página, tres bloques, y en cada uno las cuatro casillas del 4.5.1: enunciado con grupo y saber básico, coste en horas de preparación, escalón de equipamiento e indicador observable con su fecha de revisión: **lo haces solo**. Si una decisión no cabe en el escalón de papel, escribe debajo qué harías si el material no aparece: esa es la versión que se va a ejecutar. Guárdalo como **tres-decisiones.md**. **Una hora y media.**
2. **Redacta el guion de la conversación con dirección.** Media página, con el par de cifras del epígrafe 1.3 —0,65 frente a no hacer nada y 0,16 frente a otra actividad equivalente— y con lo que ese par obliga a no prometer: **lo trabajas con material público**. Debajo, lo que sí puedes acreditar de tu curso: grupos, sesiones, observaciones, fallos. Léelo en voz alta: tiene que sostenerse sin prometer retorno y sin negarlo. Guárdalo como **guion-direccion-robots.md**. **Una hora.**
3. **Monta el documento de continuidad y busca quién lo recoge.** Dos páginas con las cinco entradas del 4.5.5; después habla con quien dará el curso siguiente y con el responsable de la programación de tu equipo o departamento para que quede depositado ahí: **lo pides y esperas**. Anota la fecha de entrega y el nombre de quien lo tiene. Guárdalo como **documento-de-continuidad.md**. **Una hora.**

La tarea final optativa



1.1. Qué es, y qué no cambia

Esta tarea es **optativa** y **no condiciona el certificado**. Dicho entero y sin letra pequeña: el certificado de este curso se expide con las cuatro pruebas de bloque superadas, cada una con 12 sobre 20, y con nada más. Si no haces la tarea final optativa, tu certificado es exactamente el mismo. Nadie la va a corregir, no hay quien la recoja y no hay nota que salga de ella. La haces porque te sirve a ti, o no la haces.

Está aquí porque el curso entero apunta a un documento concreto. A lo largo de los cuatro bloques has ido dejando piezas sueltas de una misma propuesta: el saber básico localizado con la letra de la norma, el inventario de lo que hay en tu centro, el escalón desde el que arrancas, la secuencia desenchufada con sus tarjetas, el plan de subida, el protocolo de sesión con su contingencia, el reto STEAM anclado a un criterio, el encaje en el calendario y el sistema de observación. Cada pieza está en su archivo y ninguna aplicación práctica las junta. La tarea final optativa consiste en convertirlas en **la propuesta completa de pensamiento computacional para tu grupo**: un documento acabado que otro docente de tu equipo de ciclo o de tu departamento pueda conducir sin ti.

La diferencia entre las dos cosas es real. Sesenta archivos valen para que tú recuerdes lo que decidiste. Una propuesta acabada vale el lunes de la primera sesión, cuando alguien que no estuvo en tu cabeza tiene que saber qué tarjetas se reparten, qué error hay que provocar en la tercera sesión, qué pasa si el robot no empareja y qué se mira en el minuto treinta. Es también lo que se lleva a la reunión de ciclo o de departamento, y lo que el epígrafe 4.5 deja depositado para que lo herede quien venga.

Y dos fronteras que la propuesta no cruza, porque no son tuyas. La primera: aquí se decide **qué** se mira y **cuándo**; el instrumento con que se recoge —rúbrica, lista de cotejo, escala— es del **curso 07**. La segunda: el formato del proyecto largo —pregunta guía, producto público, equipos, calendario interno— es del **curso 08** y se da por conocido. Tampoco promete resultados: la evidencia que has leído en el bloque 1 no los garantiza, y esta propuesta no los escribe.

1.2. Qué se entrega

Se entrega a ti mismo. Un solo archivo, **propuesta-completa.md** o el formato que uses, guardado donde el centro guarde los documentos que llevan datos de alumnos si el tuyo los lleva, y con el alumnado identificado por código si va a salir del equipo docente. No hay dónde subirlo ni a quién mandarlo, y no hay fecha.

Se construye sobre materiales que ya produjiste. Como mínimo, estos doce: **mapa-curricular-de-mi-etapa.md** y **mi-decreto-autonomico.md**, del epígrafe 1.2; **inventario-de-cuatro-columnas.md** y **escalon-de-arranque.md**, del 1.5; **saber-producto-y-error.md**, **secuencia-desenchufada.md** y **puesta-en-comun-y-variantes.md**, del 3.1; **plan-de-subida.md**, del 3.2; **protocolo-de-sesion.md** y **plan-de-contingencia.md**, del 3.3; **encaje-en-el-calendario.md**, del 3.5; y **plan-de-observacion.md**, del 4.2. Te van a hacer falta además **definicion-operativa.md** (1.1), **saberes-sin-dispositivo.md** y **techo-del-papel.md** (1.4), **escala-de-progresion.md** (2.2), **criterio-de-escalon.md** (2.3), **puntos-de-rotura.md** y **reparto-de-papeles.md** (2.5), **decision-sobre-cuentas.md** y **normativa-dispositivos.md** (3.2), **banco-de-retos.md** (3.3), **anatomia-del-reto.md** y **reto-coste-cero.md** (3.4), **reparto-de-horas-por-area.md** y **acuerdo-de-coordinacion.md** (3.5), **matriz-producto-proceso-concepto.md** y **que-se-mira-y-cuando.md** (4.1), **sesion-observada.md** (4.2), **eleccion-de-instrumento.md** y **que-no-acredita-la-prueba.md** (4.3), **trabajo-en-aula-y-parte-individual.md** (4.4) y **tres-decisiones.md** y **documento-de-continuidad.md** (4.5). Si te faltan, hazlos primero: sin ellos esto es una propuesta de papel, y se nota en la primera página.

1.3. Extensión y apartados

Entre 2.500 y 4.000 palabras, más las tablas, las tarjetas y el calendario. Si te sale más largo, estás copiando los archivos en lugar de decidir entre ellos; si te queda mucho más corto, has resumido decisiones que otro docente necesitaría leer enteras.

Ocho apartados, en este orden.

Uno. Alcance y anclaje curricular. Qué grupo, qué etapa, qué área o áreas, qué periodo y cuántas sesiones reales. El saber básico y el criterio de evaluación seleccionados, copiados literalmente del currículo de tu comunidad, con su ciclo o curso y su anexo. La cadena completa, eslabón por eslabón: el real decreto de enseñanzas mínimas de tu etapa, tu decreto autonómico con su fecha y lo que dicen el proyecto educativo y la programación de tu equipo o departamento. Cierra con la razón escrita de por qué este saber se trabaja así, en media página, y con lo que la norma **no** te obliga a comprar.

Dos. Punto de partida. El inventario de cuatro columnas del epígrafe 1.5 —material real y su estado hoy, horas y de qué área salen, saberes que admiten anclaje, profesorado implicado con nombre— y el escalón de arranque elegido, con los tres que descartas y el dato que los descarta. Añade qué tendría que cambiar en tus columnas para subir un escalón, que es lo que evita la discusión de marzo.

Tres. Qué se puede prometer y qué no. Media página con las cifras del epígrafe 1.3 y su condición de comparación, la nota sobre lo desenchufado del 1.4 con sus límites, y la frase con la que vas a contestar en el claustro. Esta es la parte que sostiene la propuesta cuando alguien pregunte para qué sirve, y la que impide que la defiendas con algo que se cae.

Cuatro. La secuencia desenchufada. De tres a cinco sesiones escritas una por una, con el minutado, lo que hace el alumnado, lo que haces tú y lo que queda por escrito al terminar. El producto observable de cada sesión y el error que se busca, con la orden o el paso donde se provoca. Las tarjetas imprimibles en un folio por juego. Y el guion de la puesta en común, con los minutos reservados y las dos producciones que vas a comparar, más las variantes de aula sin espacio y de grupo numeroso.

Cinco. Escalones y organización de aula. El plan de subida del escenario que de verdad tienes —organización, tiempos por tramo y fallo típico con su contramedida—, la decisión escrita sobre cuentas con lo que exige el artículo 7 de la Ley Orgánica 3/2018 y quién tiene que pedir el consentimiento, y la referencia exacta de la norma de tu comunidad y del reglamento de tu centro sobre dispositivos. Si hay material compartido, el protocolo de sesión con sus tres momentos, los cuatro papeles y la rotación temporizada. Y el **plan de contingencia**, que aquí no es un anexo: es lo que garantiza que la propuesta se pueda dar el día que el aparato no responde.

Seis. El reto STEAM, si lo hay. La anatomía letra a letra, con el saber concreto que entra por cada una y lo que deja de funcionar sin ella; el criterio de evaluación que vas a poder calificar con esto; las restricciones escritas con números; la prueba del algodón pasada por escrito; y las dos versiones, la de coste cero y la ampliada, con la misma ancla y las mismas restricciones. Si tu propuesta no incluye reto, este apartado dice en dos líneas por qué y qué ocupa su sitio.

Siete. Calendario y reparto de horas. Las semanas concretas del primer trimestre en que la secuencia no la parte una evaluación, una salida ni un puente; el punto del segundo donde entraría el escalón; y lo que no vas a intentar en el tercero, con la razón. El reparto de horas entre las áreas que ya tienen el saber prescrito, sin añadir ni una hora al horario: cada sesión que anotes sustituye a una que ya trabajaba ese mismo saber de otra manera. Y el acuerdo mínimo de coordinación de una página: qué verbo y qué operación toca en cada ciclo, y qué se da por sabido al cambiar de etapa.

Ocho. Evaluación y continuidad. La matriz producto-proceso-concepto de tu actividad, con qué evidencia cada objeto y qué no permite afirmar; la declaración de qué se mira y en qué momento, lista para pegar en la programación; el plan de observación programada con su hoja

de registro; la parte de la calificación que se apoya en algo que el alumno produce por su cuenta y qué queda fuera de la nota por depender de la casa; y, si vas a pasar una prueba, cuál, para qué edad está validada y qué no acredita. Termina con las tres decisiones del curso que viene, cada una con su coste, su escalón y su indicador con fecha, y con un párrafo que diga qué no contiene esta propuesta: ni instrumentos de calificación, que son del curso 07, ni el formato del proyecto largo, que es del 08, ni la promesa de un resultado.

1.4. Pauta de autoevaluación

Como nadie lo va a corregir, el juicio lo haces tú, y con criterios, no con la sensación de haber terminado. Pasa estas diez comprobaciones sobre tu propio documento:

1. **El anclaje se sostiene al quitar la actividad.** Tapa el formato y mira el saber básico: ¿sigue en pie con otro distinto? Si al quitarlo no queda nada, tenías un juego y no una secuencia.
2. **La letra es la de la norma.** ¿Están los literales copiados del currículo, con su anexo, su ciclo y su fecha, y no reconstruidos de memoria? Las dos citas que más circulan están mal transcritas por ahí fuera; comprueba la tuya.
3. **El verbo del ciclo manda.** ¿Pide tu secuencia interpretar, modificar o crear, y coincide eso con lo que el currículo pide a esa edad? Pedir crear en primer ciclo es pedir lo que la norma no pide, por bien que quede el producto.
4. **Se puede dar sin nada.** Lee tu propuesta suponiendo que el armario está sin pilas y el aula de informática ocupada hasta junio. ¿Se da entera? Si en algún punto se para, ahí falta la versión desenchufada, y es la principal, no el plan B.
5. **El error está previsto.** ¿Hay escrito qué fallo quieres que aparezca, en qué paso y qué se hace con él? Si el error no está previsto, llega tarde o no llega, y la sesión termina con el producto bien y sin nada que discutir.
6. **Cada papel obliga a algo.** Pasa los cuatro papeles del equipo por una pregunta: ¿qué deja de salir si este no trabaja? Si la respuesta es «nada», el papel es decorativo y se abandona en dos minutos.
7. **Cada afirmación lleva su etiqueta.** Recorre el apartado tres y márcalo: qué es [E], qué es [R] y qué es [P]. Si algo aparece como «está demostrado» y no tienes la fuente delante, reescríbelo como «en mi aula funciona». Una recomendación profesional se defiende igual de bien; una cifra inventada tumba el argumento entero.
8. **Ninguna cifra va sin su condición de comparación.** ¿Dice cada número contra qué se comparó? El 0,65 y el 0,16 son el mismo estudio y la misma actividad; separarlos es prometer de más.
9. **La observación cabe en la sesión.** ¿Son tres o cuatro momentos, con su minuto, su frase y su hoja de registro? ¿Y hay una parte que cada alumno produce solo? Si toda la evidencia es del equipo, todavía no sabes quién lo domina.
10. **La prueba del compañero.** Dale el documento a alguien de tu ciclo o departamento que no conozca tu grupo y pídele que te diga qué haría en la primera sesión y qué cuando el robot no responda. Si tiene que preguntarte algo para empezar, esa parte está mal escrita.

No hay puntuación ni umbral. Si al pasar la pauta te salen tres o cuatro arreglos, está haciendo lo que tiene que hacer.

Bibliografía consolidada

Toda la identificación procede de las cuatro fichas de fuentes verificadas del curso, en **fuentes/**, comprobadas el 23 de septiembre de 2026, dentro del mes de cierre de esta edición. Las entradas siguen las normas APA, 7.ª edición, en su adaptación al español. Tras cada entrada va una línea que dice para qué sirve y en qué epígrafes. **No se añade aquí ningún número, fecha ni identificador que no conste en esas fichas:** donde una ficha no da el número, la fecha o la paginación, la entrada se queda con lo que hay y el apartado 3 explica por qué.

2.1. Normativa citada, por rango y fecha

Las dos últimas entradas son **normativa autonómica** y el manual las cita **solo como ejemplo de la dispersión** que hay entre comunidades, nunca como regla aplicable fuera de la suya.

Ley Orgánica 2/2006, de 3 de mayo, de Educación. *Boletín Oficial del Estado*, 106, de 4 de mayo de 2006. BOE-A-2006-7899, texto consolidado, en su redacción vigente tras la Ley Orgánica 3/2020. <https://www.boe.es/buscar/act.php?id=BOE-A-2006-7899> — Aporta el artículo 120, sobre autonomía pedagógica, de organización y de gestión de los centros, en los epígrafes 1.2, 1.5 y 3.5: permite decidir cuándo y cómo entra el pensamiento computacional, no si entra.

Ley 40/2015, de Régimen Jurídico del Sector Público. — Se nombra en el epígrafe 3.2 como la norma al amparo de la cual la Conferencia Sectorial de Educación es un **órgano de cooperación**, que es lo que impide llamar «norma» a sus acuerdos sobre móviles y pantallas. La ficha no recoge su fecha ni su identificador del BOE, y aquí no se le añaden; ver el apartado 3.

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, 294, de 6 de diciembre de 2018. BOE-A-2018-16673. — Aporta el artículo 7, apartados 1 y 2, sobre el consentimiento de los menores de catorce años, en el cuadro normativo del epígrafe 3.2; todo lo demás sobre plataformas y registro de menores se remite al curso 04.

Real Decreto 95/2022, de 1 de febrero, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Infantil. *Boletín Oficial del Estado*, 28, de 2 de febrero de 2022. BOE-A-2022-1654. <https://www.boe.es/buscar/act.php?id=BOE-A-2022-1654> — Área «Descubrimiento y Exploración del Entorno», competencia específica 2 y criterio de evaluación 2.5 del segundo ciclo, en los epígrafes 1.2, 1.4, 2.2 y 3.4; y el argumento de lo que **no** está en la norma —*algoritmo, robótica, actividades desenchufadas*— y del carácter no básico de los criterios y saberes del primer ciclo, por su disposición final primera.

Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. *Boletín Oficial del Estado*, 52, de 2 de marzo de 2022. BOE-A-2022-3296. <https://www.boe.es/buscar/act.php?id=BOE-A-2022-3296> — La norma más usada del manual. Anexo II, Matemáticas, bloque «D. Sentido algebraico», con la progresión de tres verbos por ciclo, y competencia específica 4 con sus criterios 4.1 y 4.2, en los epígrafes 1.2, 1.4, 1.5, 2.2, 3.1, 3.4 y 3.5; anexo II, Conocimiento del Medio Natural, Social y Cultural, competencia específica 3 y el saber del bloque B sobre recursos analógicos o digitales, en los mismos epígrafes; anexo I, descriptor CD5 del perfil de salida, en el 1.2; el artículo 11.3, con el 60 % y el 50 % de los horarios, en los epígrafes 1.2 y 3.5; y el artículo 12.1, que remite al anexo IV de horarios, en el 1.2 y el 3.5. Ver el apartado 3.

Real Decreto 217/2022, de 29 de marzo, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria. *Boletín Oficial del Estado*, 76, de 30 de marzo de 2022. BOE-A-2022-4975. <https://www.boe.es/buscar/act.php?id=BOE-A-2022-4975> — Aporta el artículo 8.3, que garantiza Tecnología y Digitalización en **uno** de los tres primeros cursos, y el artículo 9.2, que hace de Digitalización de 4.º materia de elección, en los epígrafes 1.2, 3.4 y 3.5; el

bloque de saberes «Pensamiento computacional, programación y robótica» y la competencia específica 4 de Matemáticas, en los epígrafes 1.2, 1.4, 2.2, 3.4 y 3.5; y el artículo 14.1, que remite al anexo IV, en el 1.2 y el 3.5.

Real Decreto 243/2022, de 5 de abril, por el que se establecen la ordenación y las enseñanzas mínimas del Bachillerato. *Boletín Oficial del Estado*, 82, de 6 de abril de 2022. BOE-A-2022-5521. <https://www.boe.es/buscar/act.php?id=BOE-A-2022-5521> — Establece Tecnología e Ingeniería I y II, y sostiene en los epígrafes 1.2 y 2.2 que no existe materia «TIC» estatal; y su artículo 19, que remite al anexo IV, en el 1.2 y el 3.5.

Decreto 65/2022, de 20 de julio, del Consejo de Gobierno, por el que se establecen para la Comunidad de Madrid la ordenación y el currículo de la Educación Secundaria Obligatoria. *Boletín Oficial de la Comunidad de Madrid*, de 26 de julio de 2022. https://www.bocm.es/boletin/CM_Orden_BOCM/2022/07/26/BOCM-20220726-2.PDF — Aporta, en el epígrafe 1.2 y **como ejemplo**, la optativa «Ciencias de la Computación» de 1.º y 2.º de ESO y las tres horas semanales de Tecnología y Digitalización en 2.º y dos en 3.º; y el aviso de que «Tecnología, Programación y Robótica» es denominación LOMCE que ya no figura en el decreto vigente.

Orden de 30 de mayo de 2023, por la que se desarrolla el currículo correspondiente a la etapa de Educación Secundaria Obligatoria en la Comunidad Autónoma de Andalucía. *Boletín Oficial de la Junta de Andalucía*, 104, de 2 de junio de 2023, artículo 5. <https://www.juntadeandalucia.es/boja/2023/104/36> — Aporta, en el epígrafe 1.2 y **como ejemplo**, la materia optativa propia «Computación y Robótica» en 1.º, 2.º y 3.º de ESO.

Resolución de 4 de mayo de 2022, de la Dirección General de Evaluación y Cooperación Territorial, por la que se publica el Acuerdo de la Conferencia Sectorial de Educación sobre la actualización del marco de referencia de la competencia digital docente. *Boletín Oficial del Estado*, 116, de 16 de mayo de 2022, págs. 67979 a 68026. BOE-A-2022-8042. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-8042 — Aporta, en los epígrafes 1.1 y 1.2, el literal por el que el MRCCD **excluye expresamente** el pensamiento computacional, y su naturaleza jurídica: un Acuerdo de la Conferencia Sectorial publicado por resolución, no una norma básica.

Resolución de 5 de septiembre de 2024, de la Secretaría de Estado de Educación, por la que se publica el Acuerdo de la Conferencia Sectorial de Educación de 30 de julio de 2024 sobre distribución territorial de los créditos del Programa de cooperación territorial Código Escuela 4.0 en 2024. *Boletín Oficial del Estado*, 218, de 9 de septiembre de 2024, págs. 112178 a 112187. BOE-A-2024-18091. <https://www.boe.es/buscar/doc.php?id=BOE-A-2024-18091> — Aporta, en el epígrafe 1.5, el objetivo literal del programa, su periodo de ejecución —de 1 de septiembre de 2024 a 31 de diciembre de 2026— y los 58.000.000 de euros distribuidos en 2024. Las cifras del acuerdo de Consejo de Ministros de 9 de mayo de 2023 —298,4 millones, 199 para equipamiento y 99,4 para apoyo y formación del profesorado, con 65 adicionales previstos para 2024— proceden del Portal España Digital del Gobierno de España (<https://espanadigital.gob.es/en/actualidad/acordada-una-transferencia-de-2984-millones-de-euros-para-el-programa-codigo-escuela-40>), no del BOE, y el epígrafe 1.5 las da como lo que son: financiación, no obligación curricular.

2.2. Bibliografía, por orden alfabético de autor

Agencia Española de Protección de Datos. (2018). *Informe sobre la utilización por parte de profesores y alumnos de aplicaciones que almacenan datos en nube con sistemas ajenos a las plataformas educativas. Orientaciones para centros educativos*. <https://www.aepd.es/guias/guia-orientaciones-apps-datos-alumnos.pdf> — Aporta, en el apartado 3.2.5, que «deben establecerse procedimientos que obliguen a solicitar la autorización del Centro para el uso de estas aplicaciones»: la decisión sobre qué servicio entra no es del docente a título individual. El resto se remite al curso 04.

Babazadeh, M. y Negrini, L. (2022). How is computational thinking assessed in European K-12 education? A systematic review. *International Journal of Computer Science Education in Schools*, 5(4), 1–16. <https://doi.org/10.21585/ijcses.v5i4.138> — Revisión sistemática de 26 estudios europeos con 18 instrumentos y 59 dimensiones distintas. Aporta, en los apartados 4.1.4 y 4.3.5, que la gran mayoría de esas dimensiones están más cerca de la destreza de programación que del pensamiento computacional, que no hay definición operativa común y que no existe un modelo de qué debe saber un alumno a cada edad.

Barefoot Computing. *Quick guide to computational thinking*. Programa promovido por BT junto a Computing At School, con respaldo del Departamento de Educación británico. <https://www.barefootcomputing.org/docs/default-source/at-home/quick--guide-to-computational-thinking.pdf> — Material de formación docente, no investigación. Aporta, en el apartado 2.1.1, la lista de seis conceptos y cinco enfoques con la depuración entre los segundos; y, en los apartados 2.1.2 a 2.1.4, las definiciones de reconocimiento de patrones, abstracción y algoritmo. La ficha no recoge su año de publicación, y aquí no se le pone.

Barnum, M. (2017, 21 de noviembre). *To back up claim that schools must change, DeVos cites made-up statistic about the future of work*. Chalkbeat. <https://www.chalkbeat.org/2017/11/21/21103817/to-back-up-claim-that-schools-must-change-devos-cites-made-up-statistic-about-the-future-of-work/> — Reportaje de verificación periodística. Sostiene el apartado 1.5.5: el rastreo de la cifra del 65 % se detiene en un libro de 2011 cuya autora dejó de emplearla, y el organismo internacional que más la difunde la referencia con vídeos. El manual escribe «no tiene fuente identificable», no «es falsa».

Bautista, A. (2021). STEAM education: Contributing evidence of validity and effectiveness / Educación STEAM: aportando pruebas de validez y efectividad. *Infancia y Aprendizaje / Journal for the Study of Education and Development*, 44(4), 755–768. <https://doi.org/10.1080/02103702.2021.1926678> — Presentación de un monográfico, no estudio empírico. Aporta, en el apartado 2.4.4, que la promoción del STEAM descansa sobre fundamentos empíricos relativamente débiles, dicho por quienes lo investigan.

BBC Bitesize. *The four cornerstones of computational thinking* [material divulgativo, Key Stage 3]. — Es el origen de la partición en cuatro operaciones que reproducen casi todos los materiales escolares, y la única de las listas del apartado 2.1.1 en la que **no** figura la depuración. Se cita como divulgación de un medio de comunicación, nunca como marco. Ver el apartado 3.

Bers, M. U., Flannery, L., Kazakoff, E. R. y Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, 72, 145–157. <https://doi.org/10.1016/j.compedu.2013.10.020> — Se nombra en el cuadro de alerta del apartado 2.2.1 **solo para corregir su uso**: la frase «children as young as four» está en su resumen, pero su muestra son 53 alumnos de kindergarten, de cinco y seis años.

Bers, M. U. y Resnick, M. (c. 2011). *ScratchJr* [documento de propuesta y de solicitud de financiación]. Tufts University y MIT Media Lab. <https://web.media.mit.edu/~mres/proposals/ScratchJr-draft.pdf> — Documento de diseño, no estudio. Aporta, en los apartados 2.3.3 y 2.5.2, la razón de diseño declarada por escrito: la sintaxis textual resulta poco intuitiva y frustrante para los novatos, y eso «se agrava en los niños pequeños que todavía están aprendiendo a leer», de ahí la mezcla de iconos y palabras. Es lo que sostiene la demanda lectora del entorno por bloques, que nadie ha medido.

Bocconi, S., Chiocciariello, A., Dettori, G., Ferrari, A. y Engelhardt, K. (2016). *Developing computational thinking in compulsory education: Implications for policy and practice* (EUR 28295 EN). Centro Común de Investigación (JRC), Comisión Europea. <https://doi.org/10.2791/792158> · https://publications.jrc.ec.europa.eu/repository/bitstream/JRC104188/jrc104188_computhinkreport.pdf — Fuente europea de referencia. Aporta, en los apartados 1.1.1 y 1.1.2, el deslinde entre codificar, programar y pensamiento computacional, la definición operativa de Wing de 2011 que

adopta, y el reconocimiento de que falta consenso sobre la definición; y, en el apartado 2.1.1, los seis conceptos nucleares, que son los únicos que sitúan la depuración al mismo nivel que la abstracción.

Bocconi, S., Chiocciariello, A., Kampylis, P., Dagienė, V., Wastiau, P., Engelhardt, K., Earp, J., Horvath, M. A., Jasutė, E., Malagoli, C., Masiulionytė-Dagienė, V. y Stupurienė, G. (2022). *Reviewing computational thinking in compulsory education: State of play and practices from computing education* (JRC128347). Centro Común de Investigación (JRC), Comisión Europea. <https://doi.org/10.2760/126955> · <https://publications.jrc.ec.europa.eu/repository/handle/JRC128347> — Aporta, en el apartado 1.1.3, que seis años y treinta países después los investigadores siguen sin acuerdo sobre una definición de referencia, y que pensamiento computacional, programación y codificación se usan indistintamente en los centros.

Brennan, K. y Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking* [Ponencia]. Congreso anual de la American Educational Research Association, Vancouver, Canadá, 13-17 de abril de 2012. <https://scratched.gse.harvard.edu/ct/files/AERA2012.pdf> — La fuente más usada del manual fuera del currículo. Aporta el reparto de siete conceptos, cuatro prácticas y tres perspectivas en los apartados 1.1.4 y 2.1.1; el límite declarado del análisis automático de proyectos —«this approach is entirely product-oriented, and reveals nothing about the process of developing projects»— en los apartados 4.1.2, 4.2.2 y 4.3.4; y las entrevistas basadas en artefactos y los escenarios de diseño, «process-in-action, rather than process-via-memory», en los apartados 4.1.3 y 4.2.2. Es una ponencia de congreso apoyada en treinta y una entrevistas a usuarios de Scratch, y el manual lo repite cada vez que la usa.

Chen, P., Yang, D., Metwally, A. H. S., Lavonen, J. y Wang, X. (2023). Fostering computational thinking through unplugged activities: A systematic literature review and meta-analysis. *International Journal of STEM Education*, 10, Artículo 47. <https://doi.org/10.1186/s40594-023-00434-7> — Aporta, en los apartados 1.4.2 y 3.1.4, el efecto de las actividades desenchufadas: $g = 1,028$, IC 95 % [0,641, 1,415], sobre trece estudios y dieciséis tamaños de efecto extraídos de una revisión de cuarenta y nueve. El manual da el intervalo siempre, porque es lo que impide usar la cifra para dimensionar. Los autores declaran que se limitaron a estudios en inglés y a determinadas bases de datos.

Childs, K. y Sentance, S. (2024). Investigating the impact of introducing pair programming to primary computing education on female pupils' attitudes towards computing. *International Journal of Computer Science Education in Schools*, 7(1), Artículo 218. <https://doi.org/10.21585/ijcses.v7i1.218> · <https://files.eric.ed.gov/fulltext/EJ1454640.pdf> — Se da entero cada vez que aparece, porque su resultado principal es nulo: no encontró efecto sobre la actitud de las alumnas hacia la informática —0,046 puntos en una escala de 1 a 4, $p = 0,331$ — ni sobre la intención de seguir estudiándola —4,3 puntos porcentuales, $p = 0,453$ —, en los apartados 2.5.4, 3.3.3, 4.2.4 y 4.4.4. Lo aprovechable es la práctica que documenta: la rotación de papeles cada cinco minutos con temporizador, en los apartados 3.2.4, 3.3.3, 4.2.4 y 4.4.4.

Comisión Europea, Dirección General de Educación, Juventud, Deporte y Cultura. (2019). *The 2018 International Computer and Information Literacy Study (ICILS): Main findings and implications for education policies in Europe*. Oficina de Publicaciones de la Unión Europea. ISBN 978-92-76-10363-9. <https://education.ec.europa.eu/sites/default/files/document-library-docs/icils-2018-policy-note.pdf> — Aporta, en el apartado 4.4.5, el dato que desmonta el estereotipo: en competencia informacional y digital las chicas puntuaron significativamente más alto en todos los Estados miembros participantes, de 11 puntos en Chequia y Portugal a 29 en Eslovenia y Finlandia; y en pensamiento computacional las diferencias fueron pequeñas, inconsistentes y solo significativas en Finlandia —chicas, 13 puntos— y Portugal —chicos, 16 puntos—.

Computer Science Teachers Association. (2017). *CSTA K-12 computer science standards*. <https://csteachers.org/k12standards/> — Estándares de una asociación profesional estadounidense. Aporta, en el apartado 2.2.4, la organización en niveles 1A, 1B, 2 y 3A para todo el alumnado, con el 3B reservado a estudios avanzados. Se nombra siempre como marco extranjero que no obliga en España.

CS Unplugged. Computer Science Education Research Group, University of Canterbury (Nueva Zelanda). <https://www.csunplugged.org/en/about/> — Colección de actividades libres para enseñar informática sin ordenador, «*free learning activities that teach Computer Science through engaging games and puzzles that use cards, string, crayons and lots of running around*», con licencia Creative Commons de atribución y compartir igual. Se nombra en el apartado 3.1.1 como la colección de referencia que se puede copiar y adaptar; el manual **no le atribuye ninguna lista de actividades**. Ver el apartado 3.

Csizmadia, A., Curzon, P., Dorling, M., Humphreys, S., Ng, T., Selby, C. y Woollard, J. (2015). *Computational thinking: A guide for teachers*. Computing At School (CAS) y Hodder Education. <https://www.computingschool.org.uk/media/kscbloop/computationalthinking.pdf> — Material de formación docente. Aporta, en el apartado 2.1.1, la lista de cinco conceptos —pensamiento algorítmico, descomposición, generalización, abstracción y evaluación— con la depuración fuera, como técnica; y, en los apartados 2.1.2 y 2.1.3, las definiciones de descomposición, «*in terms of their component parts*», y de abstracción, «*through reducing the unnecessary detail*».

Cuban, L. *Oversold and underused: Computers in the classroom*. Harvard University Press. Edición original en tapa dura, 2001; edición en rústica, 30 de abril de 2003, 256 páginas. ISBN 9780674011090. <https://www.hup.harvard.edu/books/9780674011090> — Se cita en el apartado 4.5.3 **como antecedente y nunca como evidencia de hoy**: trabajo de campo estadounidense de finales de los noventa, anterior a la tableta y a la banda ancha. De él se conserva la condición de partida: la implantación funciona cuando el profesorado domina la herramienta, cree que mejora el aprendizaje y conserva autonomía curricular.

Dagienė, V. y Stupurienė, G. *Informatics education based on solving attractive tasks through a contest* [Ponencia]. Recogida en el dossier de Informatics Europe, premio *Best Practices in Education* 2015 a Bebras. https://www.informatics-europe.org/images/documents/IEBPEA_2015_submission_3.pdf — Aporta, en el apartado 4.3.4, el origen del reto Bebras —2004, Lituania—, su formato de quince tareas en cuarenta minutos sin conocimientos previos de informática, su uso declarado para orientar y detectar y no para calificar, y el único dato de alcance verificado, de 2014: 33 países y más de 900.000 participantes. Las cifras de participación posteriores no se dan; ver el apartado 3.

Denner, J., Werner, L., Campe, S. y Ortiz, E. (2014). Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education*, 46(3), 277–296. <https://doi.org/10.1080/15391523.2014.888272> · ERIC EJ1090564 — Con 320 alumnos de Secundaria, aporta en los apartados 2.5.4 y 3.2.1 la ventaja de la programación por parejas y el detalle que la hace útil —quien más gana es quien menos sabe—, más el hallazgo contraintuitivo de que quien mejor se llevaba con su pareja obtuvo menos conocimiento de programación. En el apartado 3.3.3 se nombra **para decir que no vale** como respaldo de los turnos ante un robot: mide dos alumnos ante un ordenador.

Denning, P. J. (2017). Remaining trouble spots with computational thinking. *Communications of the ACM*, 60(6), 33–39. <https://doi.org/10.1145/2998438> — Sostiene el apartado 1.1.3 y el cuadro de alerta que lo cierra: las definiciones recientes se volvieron difusas y la afirmación general de que esto beneficia a todo el mundo «has never been substantiated with empirical research». El beneficio está acreditado para quien diseña computaciones.

Department for Education (Reino Unido). (2013). *National curriculum in England: Computing programmes of study*. <https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study/national-curriculum-in-england-computing->

programmes-of-study — Aporta, en el apartado 2.2.4, el reparto inglés por *key stages*: «*create and debug simple programs*» en KS1 (5-7 años), «*use sequence, selection, and repetition in programs; work with variables*» en KS2 (7-11) y dos lenguajes, «*at least one of which is textual*», en KS3 (11-14). Se usa como contraste, no como modelo a importar.

El-Hamamsy, L., Zapata-Cáceres, M., Martín-Barroso, E., Mondada, F., Dehler Zufferey, J. y Bruno, B. (2022). The competent computational thinking test: Development and validation of an unplugged computational thinking test for upper primary school. *Journal of Educational Computing Research*, 60(7), 1818–1866. <https://doi.org/10.1177/07356331221081753> — Publicación original del cCTt, en el apartado 4.3.2: 25 ítems desenchufados para 3.º y 4.º (7-9 años), $\alpha = 0,85$ sobre 1.519 alumnos de 77 aulas del cantón suizo de Vaud, y la declaración de sus autores de que mide **conceptos**, no prácticas ni perspectivas.

El-Hamamsy, L., Zapata-Cáceres, M., Marcelino, P., Bruno, B., Dehler Zufferey, J., Martín-Barroso, E. y Román-González, M. (2022). Comparing the psychometric properties of two primary school computational thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt). *Frontiers in Psychology*, 13, Artículo 1082659. <https://doi.org/10.3389/fpsyg.2022.1082659> — Es el hallazgo que decide qué prueba se pasa en 3.º y 4.º, en el apartado 4.3.2: sobre 575 alumnos de Portugal, el BCTt presenta **efecto techo** en esos cursos—13 preguntas superan el umbral máximo del índice de dificultad— y solo discrimina en el tramo de baja capacidad, mientras que el cCTt no lo presenta. Su recomendación se cita con la letra: «*For these grades, the BCTt can thus be employed as a screening mechanism to identify low ability students*».

Grover, S. y Basu, S. (2017). Measuring student learning in introductory block-based programming: Examining misconceptions of loops, variables, and Boolean logic. En *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (pp. 267–272). <https://doi.org/10.1145/3017680.3017723> — Aporta, en los apartados 2.2.3 y 2.5.1, que el alumnado de 11 a 14 años arrastra de Matemáticas la idea de que una variable es una letra que representa un número desconocido y fijo, y por eso interpreta mal las instrucciones de repetición. Ver el apartado 3.

Hermans, F. y Aivaloglou, E. (2017). To Scratch or not to Scratch? A controlled experiment comparing plugged first and unplugged first programming lessons. En *Proceedings of the 12th Workshop on Primary and Secondary Computing Education (WiPSCE '17)* (pp. 49–56). <https://doi.org/10.1145/3137065.3137072> — Experimento controlado con asignación aleatoria, treinta y cinco niños de ocho a doce años de un solo centro y ocho semanas. Aporta, en los apartados 1.4.4 y 2.3.2, que empezar por lo desenchufado no cuesta comprensión de conceptos y mejora la autoeficacia y la variedad de bloques usados por iniciativa propia. Los autores advierten de que la muestra puede no ser representativa, y el manual pone la muestra delante de la conclusión.

Hong, L. (2024). The impact of educational robots on students' computational thinking: A meta-analysis of K-12. *Education and Information Technologies*, 29(11), 13813–13838. <https://doi.org/10.1007/s10639-023-12415-y> · ERIC EJ1437292 — Aporta el SMD = 0,558 sobre 27 estudios, midiendo robots frente a enseñanza sin robots con el pensamiento computacional como único resultado, en los apartados 1.3.4 y 2.3.4. El manual **nunca escribe esa cifra sola**: va siempre junto al 0,079 no significativo de Ouyang y Xu, y sin intervalo, porque no consta. Ver el apartado 3.

Huang, W. y Looi, C.-K. (2021). A critical review of literature on «unplugged» pedagogies in K-12 computer science and computational thinking education. *Computer Science Education*, 31(1), 83–111. <https://doi.org/10.1080/08993408.2020.1789411> · ERIC EJ1288468 — Revisión de cuarenta publicaciones. Aporta, en el apartado 1.4.2, que la promesa de que lo desenchufado sea la vía de acceso equitativa a la informática no encuentra confirmación, sino zonas de desconocimiento y tensión: empezar sin aparatos resuelve un problema de material, no un problema de equidad.

IEA. (2025). *ICILS 2023 international report* (J. Fraillon y otros), capítulo 5: Student achievement in computer and information literacy and computational thinking. Springer. https://doi.org/10.1007/978-3-031-87722-3_5 — Aporta, en el apartado 4.3.6, la medida internacional de pensamiento computacional que España no hizo: 22 países y el participante de referencia, media de la escala equiparada de 483 puntos (DT 112) y un recorrido de 127 puntos entre Uruguay (421) y Taipéi Chino (548). Sin cifras por sexo; ver el apartado 3.

Instituto Nacional de Estadística. (2025, 20 de noviembre). *Encuesta sobre equipamiento y uso de tecnologías de la información y comunicación en los hogares* [Nota de prensa]. <https://www.ine.es/dyngs/Prensa/TICH2025.htm> — Aporta, en el apartado 4.4.1, que el 97,4 % de los hogares españoles disponía de acceso a internet: la cifra que impide montar el argumento de equidad sobre la conexión.

Instituto Nacional de Evaluación Educativa. (2024). *ICILS 2023. Estudio internacional de competencia digital. Informe español*. Ministerio de Educación, Formación Profesional y Deportes. <https://inee.educacion.es/2024/11/12/resultados-de-icils-2023-informe-espanol/> — Aporta, en el apartado 4.3.6, los 495 puntos de España en la escala CIL frente a 493 del promedio de la Unión Europea y 476 de media de los participantes, con 11.800 estudiantes de 2.º de ESO de 508 centros, y el hecho de que España no aplicó el módulo de pensamiento computacional; y, en el apartado 4.4.1, los 69 puntos de brecha socioeconómica frente a los 79 del promedio de la Unión Europea, los 22 puntos por tener conexión en casa (26 en la UE) y los 32 por disponer de dos o más ordenadores (39 en la UE).

Instituto Nacional de Evaluación Educativa. (2026, 16 de enero). *ICILS 2028: la primera medición del pensamiento computacional en España* [Entrada de blog]. <https://inee.educacion.es/2026/01/16/icils-2028-la-primera-medicion-del-pensamiento-computacional-en-espana/> — Aporta, en los apartados 4.3.6 y 4.5.5, que España no tomó parte en el módulo opcional de pensamiento computacional de ICILS 2023 y que la primera medición española será en 2028, con alumnado que hoy cursa 6.º de Primaria. Es argumento de calendario, no de resultados.

Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado. (2022). *Escuela de Pensamiento Computacional e Inteligencia Artificial 20/21. Enfoques y propuestas para su aplicación en el aula. Resultados de la investigación*. Ministerio de Educación y Formación Profesional. NIPO 847-21-260-X. https://code.intef.es/wp-content/uploads/2022/04/Informe-EPCIA20_21.pdf — El único dato español del manual. Aporta, en los apartados 1.3.5, 1.4.2, 3.1.4 y 3.5.4, el Nivel I de la edición 20/21 con actividades desenchufadas: 18 comunidades y ciudades autónomas, 119 docentes, 2.580 estudiantes —1.824 experimentales y 756 de control—, más del 80 % de las propuestas en cinco o seis sesiones, y los efectos $d = 0,315$ en Infantil de cinco años, 0,156 en 1.º, 0,185 en 2.º y 0,228 en 3.º de Primaria. Los autores declaran tres limitaciones que el manual repite cada vez: diseño solo post-test, grupos no aleatorios y sin prueba previa.

K-12 computer science framework. (2016). Association for Computing Machinery, Code.org, Computer Science Teachers Association, Cyber Innovation Center y National Math and Science Initiative. <https://k12cs.org/> — Aporta, en el apartado 2.2.4, su reparto en cuatro bandas —K-2, 3-5, 6-8 y 9-12— y su propia declaración de que es guía conceptual y no estándares ni currículo. Ver el apartado 3.

Kirçali, A. Ç. y Özdener, N. (2023). A comparison of plugged and unplugged tools in teaching algorithms at the K-12 level for computational thinking skills. *Technology, Knowledge and Learning*, 28(4), 1485–1513. <https://doi.org/10.1007/s10758-021-09585-4> · ERIC EJ1394416 — Con 109 alumnos de sexto durante seis semanas en tres condiciones, aporta en el apartado 1.4.4 que el único grupo que mejoró respecto de sí mismo fue el desenchufado y que entre los tres grupos no aparecieron diferencias significativas. Ni gana el papel ni da igual: con esa muestra, el estudio no separa las condiciones.

Li, Z. y Oon, P. T. (2024). The transfer effect of computational thinking (CT)-STEM: A systematic literature review and meta-analysis. *International Journal of STEM Education*, 11, Artículo 44. <https://doi.org/10.1186/s40594-024-00498-z> — Revisión sistemática con metaanálisis sobre 37 estudios, 7.832 estudiantes y 96 tamaños de efecto. Aporta, en el apartado 1.3.3, que sus efectos se describen como **moderados** y, sobre todo, sus moderadores: cuatro significativos para el efecto global y la transferencia cercana —nivel educativo, tamaño muestral, estrategia instruccional y duración— y solo dos para la lejana. **El manual no le pone ninguna cifra**; ver el apartado 3.

Michaeli, T. y Romeike, R. (2019). Improving debugging skills in the classroom: The effects of teaching a systematic debugging process. En *Proceedings of the 14th Workshop in Primary and Secondary Computing Education (WiPSCE '19)*. ACM. <https://doi.org/10.1145/3361721.3361724> — Es el único ensayo con grupo de control sobre enseñanza explícita de un procedimiento de depuración en un aula escolar que este curso ha podido comprobar, y se da con su tamaño en el apartado 2.1.5: 28 alumnos en el contraste principal, de 15 a 16 años, con una intervención de unos diez minutos dentro de una sesión de noventa y en lenguaje textual. Da efectos significativos y no autoriza a generalizar.

Moreno-León, J., Robles, G. y Román-González, M. (2015). Dr. Scratch: análisis automático de proyectos Scratch para evaluar y fomentar el pensamiento computacional. *RED. Revista de Educación a Distancia*, (46), Artículo 10. <https://revistas.um.es/red/article/view/240251> — Publicación original de la herramienta, en el apartado 4.3.4: siete dimensiones puntuadas de 0 a 21 y la declaración de sus autores de que es un instrumento **formativo**.

Moreno-León, J., Robles, G. y Román-González, M. (2016). *Comparing computational thinking development assessment scores with software complexity metrics* [Ponencia]. IEEE Global Engineering Education Conference (EDUCON 2016), Abu Dabi. <https://www.researchgate.net/publication/296148316> — Aporta, en los apartados 4.1.2 y 4.3.4, la validación de la puntuación de Dr. Scratch contra métricas clásicas de complejidad del software sobre 95 proyectos: complejidad ciclomática de McCabe ($r^2 = 0,385$), vocabulario ($r^2 = 0,461$) y longitud ($r^2 = 0,319$) de Halstead. De ahí que la puntuación se comporte como una medida de complejidad del código.

OCDE. (2014). *Does homework perpetuate inequities in education?* (PISA in Focus n.º 46). OECD Publishing. <https://doi.org/10.1787/5jxrhqhtx2xt-en> — Se cita **como analogía** en el apartado 4.4.2: el alumnado de entorno favorecido dedica más tiempo al trabajo fuera del aula. Es de deberes, no de robótica, y el manual lo dice.

Ofqual. (2019). *Future assessment arrangements for GCSE computer science* (Ofqual/18/6439). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/771555/Consultation_Future_assessment_arrangements_for_GCSE_computer_science_Ofqual_18_6439.pdf — El precedente más directo que existe, en el apartado 4.4.2: el regulador inglés retiró de la nota el proyecto de programación del título de secundaria de Informática porque no podía garantizar de quién era el trabajo, y lo mantuvo obligatorio, veinte horas lectivas, sin calificación. Con su letra pequeña: el motivo declarado fue la filtración de soluciones, no la desigualdad de recursos.

Oliveira, A. L. S., Andrade, W. L., Serey, D. y Melo, M. R. A. (2025). Quantifying computational thinking skills: An exploratory study on Bebras tasks. *Journal of the Brazilian Computer Society*, 31(1). <https://journals-sol.sbc.org.br/index.php/jbcs/article/view/3893> — Análisis con teoría de respuesta al ítem que aporta, en el apartado 4.3.4, la cautela sobre Bebras: la dificultad asignada a priori a cada tarea acierta solo entre el 53 % y el 58 % de las veces, y las tareas se concentran de fácil a medio, de modo que el instrumento discrimina mal arriba.

Ouyang, F. y Xu, W. (2024). The effects of educational robotics in STEM education: A multilevel meta-analysis. *International Journal of STEM Education*, 11, Artículo 7. <https://doi.org/10.1186/s40594-024-00469-4> · ERIC EJ1410342 — Sobre 21 estudios y 30 tamaños de efecto, aporta en los apartados 1.3.4 y 2.3.4 el efecto global $g = 0,488$ [0,094, 0,882], el rendimiento de aprendizaje

0,665, las actitudes 0,497 [0,023, 1,017] y —lo decisivo— $g = 0,079$, $p = 0,434$, IC 95 % [-0,119, 0,277] sobre pensamiento computacional: **no significativo**. De ocho moderadores probados solo resultó significativa la disciplina, y la condición del grupo de control **no**, de modo que ese argumento se apoya en Scherer y no aquí.

Perignat, E. y Katz-Buonincontro, J. (2019). STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 31, 31–43. <https://doi.org/10.1016/j.tsc.2018.10.002> — Revisión integradora sobre 44 artículos publicados entre 2007 y 2018. Aporta, en el apartado 2.4.1, «a myriad of definitions of the STEAM concept in general, a variety of interpretations for the “A” in STEAM» y los cinco modelos de integración; y, en el apartado 2.4.4, «an overall lack of reported learning outcomes in the areas of creativity, problem-solving, and arts education» y «Creativity is rarely, if ever, measured or assessed as part of STEAM education».

Relkin, E., de Ruiter, L. y Bers, M. U. (2020). TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education. *Journal of Science Education and Technology*, 29(4), 482–498. <https://doi.org/10.1007/s10956-020-09831-x> — Aporta, en el apartado 4.3.3, la prueba desenchufada para el tramo de 5 a 9 años: 15 ítems puntuables más dos de práctica, duración media real de 13,40 minutos, 768 alumnos de 1.º y 2.º de Primaria, α de Cronbach 0,68 —bajo—, posible efecto techo con un 8 % en la puntuación máxima y, en su publicación original, **sin validar en Infantil**. Ver el apartado 3.

Robins, A., Rountree, J. y Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, 13(2), 137–172. <https://doi.org/10.1076/csed.13.2.137.14200> — Revisión de referencia sobre aprendizaje de la programación, con literatura universitaria, y así se nombra. Aporta, en el apartado 2.5.1, la distinción que ordena la contramedida: el principiante falla más por falta de estrategia que de conocimiento. No se entrecomilla; ver el apartado 3.

Román-González, M. (2015). *Computational Thinking Test: Design guidelines and content validation* [Ponencia]. EDULEARN15, 7th International Conference on Education and New Learning Technologies, Barcelona, 6-8 de julio de 2015, pp. 2436–2444. ISBN 978-84-606-8243-1. — Publicación original del CTt, en los apartados 4.1.4 y 4.3.1: 28 ítems de opción múltiple, 45 minutos, aplicación colectiva, sobre siete conceptos, diseñado para alumnado español de 12 y 13 años. Lo más útil es lo que declara su autor: el formato cerrado hace que el test «only measures CT at their lower levels of complexity (“Recognize” and “Understand”)», y él mismo se pregunta si no será «“just” an AT-test (Algorithm Thinking Test)».

Román-González, M., Moreno-León, J. y Robles, G. (2017). *Complementary tools for computational thinking assessment* [Ponencia]. International Conference on Computational Thinking Education 2017 (CTE2017), Hong Kong. — Aporta, en el apartado 4.3.4, cuánto se complementan los tres instrumentos: $r = 0,52$ entre el CTt y un conjunto de tareas Bebras ($n = 179$), y $r = 0,44$ predictivo y $r = 0,53$ concurrente entre el CTt y Dr. Scratch ($n = 71$). Y, en el apartado 4.1.2, la clasificación de Dr. Scratch como instrumento de evaluación **formativa**.

Román-González, M., Pérez-González, J. C. y Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*, 72, 678–691. <https://doi.org/10.1016/j.chb.2016.08.047> — Sobre 1.251 estudiantes de 5.º de Primaria a 4.º de ESO, aporta en los apartados 4.1.4 y 4.3.5 las tres correlaciones que dejan abierta la pregunta de cuánto de lo que se mide es específico de esta materia: $r = 0,67$ con resolución de problemas, $r = 0,44$ con razonamiento y $r = 0,44$ con capacidad espacial. Ver el apartado 3.

Rønning, M. (2011). Who benefits from homework assignments? *Economics of Education Review*, 30(1), 55–64. <https://doi.org/10.1016/j.econedurev.2010.07.001> — Se cita **como analogía** en el apartado 4.4.2: en alumnado de Primaria de los Países Bajos, la diferencia de resultados es mayor en las clases donde se mandan deberes a todos, porque el alumnado desfavorecido recibe menos ayuda en casa. Es de deberes, no de robótica.

Scherer, R., Siddiq, F. y Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology*, 111(5), 764–792. <https://doi.org/10.1037/edu0000314> · ERIC EJ1220317 — La fuente que ordena el curso entero. Metaanálisis de tres niveles y efectos aleatorios sobre 105 estudios y 539 tamaños de efecto. Aporta, en el apartado 1.3.1, $g = 0,49$ [0,37, 0,61] global, transferencia cercana 0,75 [0,39, 1,11] y lejana 0,47 [0,35, 0,59], con pensamiento creativo 0,73, destrezas matemáticas 0,57, metacognición 0,44, destrezas espaciales 0,37 y razonamiento 0,37, rendimiento escolar 0,28 [0,14, 0,42] y competencia lectora en un efecto indistinguible de cero (IC 95 % [-0,12, 0,08]); y, en el apartado 1.3.2, el par que atraviesa el manual: 0,65 [0,51, 0,78] frente a un control que no hace nada y 0,16 [0,01, 0,33] frente a otra actividad equivalente, más el sesgo de publicación medido —0,60 [0,45, 0,75] publicado frente a 0,34 [0,15, 0,52] gris—. Se reutiliza en los apartados 2.3.4 y 4.5.4. Ver el apartado 3.

Scherer, R., Siddiq, F. y Sánchez-Scherer, B. (2021). Some evidence on the cognitive benefits of learning to code. *Frontiers in Psychology*, 12, Artículo 559424. <https://doi.org/10.3389/fpsyg.2021.559424> — **Artículo de opinión**, no estudio empírico, y así se cita. Aporta, en el apartado 1.3.2, dos precisiones de los mismos autores: la transferencia no ocurre sola, requiere un diseño instruccional orientado a ella, y con los diseños no aleatorizados que dominan esta literatura no caben conclusiones estrictamente causales. De ahí que el manual escriba «se asoció con» y no «causó». Ver el apartado 3.

Selby, C. C. y Woollard, J. (2013). *Computational thinking: The developing definition* [Project Report]. University of Southampton, ePrints Soton 356481. <https://eprints.soton.ac.uk/356481/> — **Informe de proyecto universitario, no artículo revisado por pares**, y así hay que citarlo. Aporta, en el apartado 1.1.4, su lista de cinco procesos —abstracción, descomposición, diseño algorítmico, evaluación y generalización— sin el reconocimiento de patrones como operación aparte. Ver el apartado 3.

Shute, V. J., Sun, C. y Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142–158. <https://doi.org/10.1016/j.edurev.2017.09.003> — Es la definición que este curso adopta, en el apartado 1.1.4, diciendo que la elección es suya y no un consenso del campo: el fundamento conceptual necesario para resolver problemas de forma eficaz y eficiente, algorítmicamente, **con o sin ayuda de ordenadores**, con soluciones reutilizables en contextos distintos. Se elige por ese inciso y porque mete la depuración dentro del constructo.

Staker, H. y Horn, M. B. (2012). *Classifying K-12 blended learning*. Clayton Christensen Institute (entonces Innosight Institute). <https://www.christenseninstitute.org/publication/classifying-k-12-blended-learning-2/> — Es una **taxonomía que clasifica lo que existe**, no una medición de eficacia, y así se presenta en los apartados 1.5.2 y 3.2.1, donde el manual la nombra por la institución: de ahí procede el vocabulario de la rotación por estaciones que sostiene el escenario del rincón. Ver el apartado 3.

Südkamp, A., Kaiser, J. y Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627> — Metaanálisis sobre 75 estudios que aporta, en el apartado 4.2.1, la correlación media de $r = 0,63$ entre el juicio del docente y el rendimiento medido con prueba estandarizada, y el moderador que da la regla de trabajo: el juicio **informado** se corresponde mejor con el rendimiento real. No es un estudio sobre robótica ni sobre programación escolar, y el manual lo dice.

Sullivan, A. y Bers, M. U. (2015). Robotics in the early childhood classroom: Learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *International Journal of Technology and Design Education*, 25(3). <https://doi.org/10.1007/s10798-015-9304-5> — Es el dato verificable de los cuatro años, y se da siempre con su muestra. Aporta, en los apartados 2.2.1 y 2.5.3, que el grupo de Pre-K —quince niños— dominó la secuencia corta con una media de 5,2 sobre 6, bajó a 3,4 sobre 6 en la secuencia larga y **no llegó a bucles, condicionales ni sensores**; en el

apartado 2.2.2, que Infantil, 1.º y 2.º rindieron de forma equivalente en las tareas avanzadas, con la única diferencia significativa por curso en la secuencia larga ($p = 0,00023$); y, en el apartado 3.3.2, que el alumnado de cuatro años necesitó más apoyo adulto y más repetición. Es un estudio sin grupo de control: da techo, no eficacia.

Sun, C., Yang, S. y Becker, B. J. (2024). Debugging in computational thinking: A meta-analysis on the effects of interventions on debugging skills. *Journal of Educational Computing Research*, 62(4). <https://doi.org/10.1177/07356331241227793> — Sobre 62 tamaños de efecto procedentes de 18 artículos, aporta en los apartados 2.1.5 y 4.2.3 el efecto medio de las intervenciones sobre la destreza de depuración: $g = 0,64$, IC 95 % [0,32, 0,96], con la instrucción sistemática en $g = 0,93$, los enfoques basados en juego en $g = 0,94$ y los mensajes de error mejorados en $g = 0,04$, efecto nulo. Los autores declaran heterogeneidad alta y posible sesgo de publicación, y el manual lo repite.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4 — Se nombra en el apartado 2.5.1 **diciendo expresamente lo que se está haciendo con él**: sus experimentos son de geometría y álgebra, no de programación, y no existe ninguna medición verificada de carga cognitiva en una sesión escolar de robótica. El manual aplica un mecanismo general a un caso que nadie ha medido, y lo declara. El estudio de la memoria de trabajo es del curso 17.

Wang, J., Fan, W. y Yu, J. (2025). Comparing the effects of unplugged activities and plugged activities on the development of students' computational thinking: A meta-analysis. *Educational Technology Research and Development*, 73(5), 3247–3267. <https://doi.org/10.1007/s11423-025-10522-w> — Metaanálisis comparativo sobre 37 estudios publicados entre enero de 2006 y marzo de 2025. Aporta las dos cifras que **no se restan**, cada una medida contra su propio grupo de control —enchufado $g = 0,606$, desenchufado $g = 0,501$ —, en el apartado 1.4.3; y, sobre todo, los moderadores que ordenan la programación: equivalencia en Educación Primaria, en materias STEM y en intervenciones de más de cuatro semanas, ventaja de lo enchufado en Secundaria y en intervenciones cortas, y ventaja de lo desenchufado en aprendizaje basado en juego, en los apartados 1.4.3, 2.3.5, 3.1.4, 3.2.4 y 3.5.4. Ver el apartado 3.

Weintrop, D. y Wilensky, U. (2017). Comparing block-based and text-based programming in high school computer science classrooms. *ACM Transactions on Computing Education*, 18(1), Artículo 3, 1–25. <https://doi.org/10.1145/3089799> — Estudio cuasiexperimental de cinco semanas con 60 alumnos, dos grupos de un mismo profesor en un único instituto de acceso selectivo. Es el origen de la ventaja de los bloques sobre el texto a las cinco semanas —66,6 % frente a 58,8 %— que los apartados 2.2.3 y 2.3.3 usan, y de la diferencia que no se igualó: el interés por seguir.

Weintrop, D. y Wilensky, U. (2019). Transitioning from introductory block-based and text-based environments to professional programming languages in high school computer science classrooms. *Computers & Education*, 142, Artículo 103646. <https://doi.org/10.1016/j.compedu.2019.103646> — Continuación del anterior con el mismo alumnado, cinco semanas más diez en Java. Aporta el dato decisivo de los apartados 2.2.3 y 2.3.3: la ventaja de los bloques **desapareció** a las quince semanas —64,9 % frente a 64,7 %—, sin diferencias en prácticas de programación. Con 60 alumnos de un centro selectivo, es una pista y no una ley, y así se escribe.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215> — Es la columna de opinión que pone el término en circulación, y el manual dice qué es antes de citarla. Aporta, en el apartado 1.1.2, «*involves solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science*», «*a fundamental skill for everyone, not just for computer scientists*» y el par que más falta hace en un claustro: «*Conceptualizing, not programming*». En el cuadro de alerta del mismo epígrafe se nombra para advertir de que la frase de la alfabetización del siglo XXI procede de aquí, y que esto es un manifiesto de dos páginas y media, no un resultado.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A*, 366(1881), 3717–3725. <https://doi.org/10.1098/rsta.2008.0118> — Aporta, en el apartado 1.1.2, la reformulación acotada: «*The essence of computational thinking is abstraction*» y «*Computing is the automation of our abstractions*».

Winner, E., Goldstein, T. R. y Vincent-Lancrin, S. (2013). *Art for art's sake? The impact of arts education*. OECD Publishing. ISBN 9789264180789. <https://doi.org/10.1787/9789264180789-en> — Sostiene el apartado 2.4.3 entero y se recuerda en el 3.4.3: no hay evidencia de que la formación en artes plásticas mejore destrezas verbales o matemáticas, ni de impacto causal de la educación musical sobre las puntuaciones de matemáticas, ni de vínculo entre formación teatral y destrezas académicas generales, y los estudios experimentales multiartísticos no muestran impacto significativo sobre el rendimiento. Y la excepción acotada que el manual da igual que lo anterior: la dramatización de relatos en el aula **sí** refuerza las destrezas verbales.

Zapata-Cáceres, M., Martín-Barroso, E. y Román-González, M. (2020). Computational thinking test for beginners: Design and content validation. En *2020 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1905–1914). <https://doi.org/10.1109/EDUCON45650.2020.9125368> — Publicación original del BCTt, en el apartado 4.3.2: 25 ítems de opción múltiple, desenchufados, para 1.º a 6.º de Primaria, con $\alpha = 0,82$ sobre 200 alumnos de validación. Su efecto techo en 3.º y 4.º es un hallazgo posterior, de El-Hamamsy y otras (2022).

Referencias citadas con reserva

Estas fuentes se han usado con cautela, y aquí se dice cuál y por qué. Ninguna se identifica con un número, una fecha o una paginación que no consten en las fichas.

La definición de Wing de 2011. El apartado 1.1.2 la da entera —«los procesos de pensamiento que intervienen en formular problemas y sus soluciones de modo que la solución se represente de una forma que un agente de procesamiento de información pueda ejecutar»— **a través del informe del JRC de 2016**, que es quien la adopta y la atribuye a Wing (2011), y el epígrafe lo escribe así. La nota original de la CMU no se ha podido abrir y por eso **no lleva entrada propia** en la bibliografía.

Los anexos II de los cuatro reales decretos. El acceso directo al BOE estuvo bloqueado durante la verificación —el texto consolidado se trunca antes de los anexos de áreas y materias—, de modo que los literales se cotejaron contra **dos fuentes oficiales del Ministerio de Educación, Formación Profesional y Deportes, concordantes entre sí**: el portal INTEF de Código Escuela 4.0, que reproduce los saberes básicos por área, bloque y ciclo, y el extracto de Conocimiento del Medio de 4.º publicado por un centro del Ministerio en el exterior. **El cotejo se completó y los seis literales quedaron confirmados palabra por palabra**, de modo que los apartados 1.5.1 y 3.4.2 los citan entre comillas latinas identificando el real decreto, el anexo, el área, el bloque y el ciclo. Lo único que el manual no hace es **atribuirles número de página del BOE**, porque la página no se ha visto. Por la misma razón escribe el **título** del bloque de Tecnología y Digitalización y no su letra: no se ha podido confirmar si el BOE lo rotula «C» o «3».

El anexo IV de horarios: la única cifra que este manual no da. Es la excepción del cotejo. No hay fuente oficial cotejada para las horas por área y ciclo, y por eso **el manual no escribe ninguna**: las que circulaban —del orden de 180 horas de Matemáticas por ciclo y 160 de Conocimiento del Medio— proceden de la ficha de España de Eurydice, no del texto oficial, y se han retirado. El apartado 3.5.1 dice lo que sí consta: que el anexo IV expresa el horario en horas por área y ciclo, que cada administración educativa lo concreta y que el número hay que tomarlo de ahí, no de este manual. **De ESO tampoco se da ninguna cifra.** Lo que sí está verificado, y por eso se escribe sin reserva, es el artículo que remite al anexo IV y el reparto del 60 % y el 50 % del artículo 11.3.

Ley 40/2015. La ficha la nombra solo como la norma al amparo de la cual la Conferencia Sectorial de Educación es un órgano de cooperación. No recoge su fecha ni su identificador del BOE, y aquí no se le añaden. El apartado 3.2.6 la usa únicamente para eso: para decir que un acuerdo de ese órgano no es una prohibición estatal. Y tampoco escribe «acuerdo de la Conferencia Sectorial sobre pantallas» como si fuera norma: lo verificado es la reunión de 13 de diciembre de 2023, el grupo de trabajo reunido el 31 de enero de 2024 y la propuesta que el Ministerio puso sobre la mesa; **no consta ningún acuerdo formal publicado en el BOE** con texto articulado, ni para 2025 ni para 2026.

Li y Oon (2024): efectos sin cifra. El resumen oficial describe los efectos como «moderate» y no da números, y el texto completo se cortó antes de la sección de resultados en las tres consultas. El apartado 1.3.3 escribe «moderados» —que es literal— y **no pone ningún número a esos efectos**, ni toma prestados los de Scherer. Lo verificado, y lo que se usa, son los moderadores.

Hong (2024): el 0,558 sin intervalo. Ni el resumen oficial ni la ficha de ERIC incluyen el IC 95 % del SMD, y el texto completo es de pago. Por eso la cifra se escribe sin intervalo y **nunca sola**: va siempre junto al 0,079 no significativo de Ouyang y Xu, y el cuadro de alerta del apartado 1.3.4 lo dice expresamente.

Wang, Fan y Yu (2025): limitaciones no verificadas. Están verificados el diseño, el número de estudios, las fechas, las dos cifras y la tabla de moderadores. **No se ha verificado ninguna limitación declarada**, ni el análisis de sesgo de publicación ni la heterogeneidad, y por eso el manual no afirma que el estudio controlara el sesgo de publicación.

Scherer y otros (2019): el signo del efecto en lectura. La tabla consultada imprime un punto estimado que queda fuera de su propio intervalo, y el artículo de opinión de 2021 de los mismos autores apunta a que el valor es negativo. **El manual no escribe el signo:** da el intervalo $[-0,12, 0,08]$ y la frase de los autores, que no encontraron apoyo para una transferencia positiva a la lectura. De la sección de limitaciones solo se han verificado dos, y solo esas dos se citan.

Selby y Woollard (2013). El PDF íntegro devuelve 403. Están verificados en la ficha del repositorio la autoría, el año, el tipo de publicación y la lista de cinco procesos; **el manual no entrecomilla ninguna otra frase** de este informe.

Robins, Rountree y Rountree (2003) y Grover y Basu (2017). De la primera no se ha podido obtener texto literal fiable, y por eso el apartado 2.5.1 **la parafrasea sin comillas**. De la segunda, la identificación bibliográfica está completa, pero el hallazgo sobre la concepción matemática de «variable» procede de reseña académica pública y no del artículo: por eso el manual lo describe y **no da cifras** de ese estudio.

La lista «de la BBC». El servidor de la BBC devuelve bloqueo al lector automático y **no se ha obtenido cita literal de la página original**. El apartado 2.1.1 la describe y la atribuye a la divulgación —y a la literatura académica que la reproduce—, sin entrecomillar nada, y la usa para decir lo único que importa: que la partición en cuatro operaciones que reproducen casi todos los materiales escolares procede de ahí y no de un experimento, y que en ella la depuración no aparece. Por eso su entrada va sin año.

Barefoot Computing. Material de formación docente sin año de publicación en la ficha. Se cita por su contenido, nunca como hallazgo de investigación.

CS Unplugged. La página consultada describe la colección y su licencia, pero **no enumera las actividades**. Por eso el apartado 3.1.1 declara que su repertorio de siete formatos es **repertorio profesional de uso común** y no el catálogo de nadie, y remite a CS Unplugged solo como la colección de referencia con licencia libre.

Staker y Horn (2012). Están confirmadas la autoría, la fecha, la naturaleza taxonómica y los nombres de las cuatro subcategorías; **no las definiciones textuales**. Los apartados 1.5.2 y 3.2.1 describen el modelo y no lo citan, y dicen que es una clasificación de lo que existe y no una medición de eficacia.

El K-12 Computer Science Framework y los estándares CSTA. El sitio del primero no declara el año en texto; se escribe 2016 sobre la base de las rutas de archivo y de la literatura que lo cita. Del segundo, el visor de estándares devuelve 403: están confirmadas la estructura de niveles y que el 3B es para estudios avanzados, **no las edades ni el texto literal por nivel**, y el apartado 2.2.4 no los entrecomilla.

Denner y otros (2014). Verificados el diseño, la muestra de 320 alumnos y el sentido de los resultados; **sin tamaños de efecto, p ni medias**, que por eso no aparecen en los apartados 2.5.4 ni 3.2.1.

Relkin, de Ruiter y Bers (2020). Existe una adaptación posterior de TechCheck para Infantil cuyas edades, muestra y fiabilidad no se han leído. El apartado 4.3.3 dice lo que está verificado —que la prueba **no estaba validada en Infantil** en su publicación original— y no da dato alguno de esa adaptación.

Román-González, Pérez-González y Jiménez-Fernández (2017). El texto completo está tras muro de pago y solo se ha leído el resumen, del que proceden la muestra de 1.251, los cursos y las tres correlaciones. El α del CTt, sus 28 ítems y sus 45 minutos están verificados contra la ponencia EDULEARN15 del mismo autor, no contra este artículo. Las **diferencias por sexo en el CTt no están verificadas y no se escriben**.

ICILS 2023 y las cifras por sexo. El capítulo 5 del informe internacional declara que hay patrones de rendimiento por sexo distintos entre competencia digital y pensamiento computacional, pero **las tablas con esas cifras no se sirven** en la versión consultada. Por eso el apartado 4.4.5 no escribe ninguna cifra de sexo de la escala de pensamiento computacional de ICILS 2023 y apoya el argumento en el informe de la Comisión Europea sobre ICILS 2018, que es explícito. Y una precaución de lectura que el manual respeta: los 19 puntos a favor de las chicas del informe español son de la escala CIL, **no** de pensamiento computacional.

Bebras: las cifras de alcance. El dato verificado —33 países y más de 900.000 participantes— es **de 2014**, y así se fecha en el apartado 4.3.4. Las cifras de participación posteriores que circulan no se han verificado y no se usan.

OCDE (2015), *Students, Computers and Learning*. La frase que se le suele atribuir sobre la ausencia de mejoras apreciables solo se ha verificado en fuente secundaria. El manual **no la escribe**, y tampoco escribe ya las cifras del informe internacional de 2023 que antes la acompañaban: por qué el valor añadido de la tecnología digital está mal documentado es del **curso 04**, al que el apartado 4.5.3 remite en una frase, y el argumento lo sostiene allí el par $g = 0,65 / g = 0,16$ de Scherer, Siddiq y Sánchez Viveros (2019).

Lo que no tiene evidencia y se escribe como tal. Hay tres afirmaciones centrales del manual que **no descansan en ningún estudio**, y los epígrafes lo dicen en su propio texto: que sin una máquina que ejecute literalmente lo escrito la depuración auténtica no tiene dónde ocurrir (apartados 1.4.5 y 2.3.2); que no existe medición del número de robots por aula, de la duración del turno ni del tiempo que se va en baterías y emparejamientos (apartados 2.5.5, 3.3.1 y 3.3.4); y que evaluar el proceso dé notas más justas que evaluar el producto, o que las cuatro decisiones de calificación del 4.4 corrijan la desigualdad (apartados 4.1.5, 4.2.4, 4.4.2 y 4.4.4). Las tres van marcadas **[P]**. Tampoco existe **ningún metaanálisis de eficacia de la educación STEAM**, y por eso en este manual no hay ni una cifra de eficacia del STEAM (apartado 2.4.4). Ni nada verificable sobre robótica escolar y vocaciones técnicas: el manual no lo escribe en ninguna forma, y para la orientación hacia lo técnico remite al curso 16.