

Evidencia disponible sobre el trabajo por proyectos y factores que la condicionan

Saber exactamente qué se puede afirmar sobre el ABP y qué no, y traducir esa evidencia a decisiones concretas de andamiaje, agrupamiento y duración.

- 1 Resultados de los ensayos y metaanálisis disponibles
- 2 Variables moderadoras: duración, tamaño del equipo, etapa educativa y materia
- 3 Efectos del proyecto sin andamiaje sobre el alumnado con menor nivel previo
- 4 Enseñanza previa necesaria y retirada progresiva de la ayuda
- 5 Distinción entre trabajo en grupo y aprendizaje cooperativo

2.1. Resultados de los ensayos y metaanálisis disponibles

Qué dicen los ensayos y los metaanálisis, y por qué no dicen lo mismo.

En algún claustro alguien dirá que está demostrado que el trabajo por proyectos mejora los resultados, y otro contestará que en Inglaterra se probó y salió mal. Los dos tendrán un estudio detrás, y los dos estarán citando la mitad. Hay resultados en las dos direcciones, y casi nunca viajan con su letra pequeña: qué se probó exactamente, frente a qué, qué se midió y qué se perdió por el camino. Leídos enteros, no dan la razón ni al folleto ni al escéptico.

2.1.1. Tres ensayos favorables, y qué probaron de verdad

Los resultados más sólidos a favor proceden de tres ensayos aleatorizados financiados por Lucas Education Research y publicados entre 2020 y 2023. Son evidencia empírica directa, con asignación al azar.

El primero, Knowledge in Action, se aplicó en los cursos 2016-17 y 2017-18 en cinco distritos urbanos de Estados Unidos, en asignaturas de secundaria superior con examen externo AP (*Advanced Placement*) (Saavedra y otros, 2021, 2022). En el primer año, con 68 centros y 3.645 alumnos, la proporción de aprobados subió unos cuatro puntos porcentuales si se cuenta a todo el alumnado matriculado (del 31,1 al 35,1 %) y 7,6 si se cuenta solo a quien se presentó (del 37,2 al 44,8 %). El efecto también fue significativo en el alumnado de renta baja. El informe declara dos limitaciones: el segundo año no fue concluyente, y la pérdida de centros superó los umbrales de calidad del What Works Clearinghouse.

El segundo, ML-PBL, se hizo con alumnado de tercero de Primaria en 46 centros de Míchigan (Krajcik y otros, 2023). El informe técnico de 2021 traduce su efecto, 0,266 desviaciones típicas, en ocho puntos porcentuales en la prueba externa de ciencias del estado, y el efecto se mantuvo con independencia del nivel lector y del sexo. El tercero, Project PLACE, trabajó con 48 docentes de segundo de Primaria en distritos de bajo nivel socioeconómico (Duke y otros, 2021): un tamaño del efecto de 0,48 en ciencias sociales y de 0,18 en lectura informativa, sin efecto significativo en escritura ni en motivación.

Ninguno probó «el proyecto» como formato suelto. Knowledge in Action dio cinco unidades completas con materiales y ocho días de formación docente. ML-PBL, cuatro unidades con materiales, unas 45 horas de formación y evaluaciones tras cada unidad. Project PLACE, cuatro unidades con planes de sesión y visitas de acompañamiento, y una mayor fidelidad a esos planes se asoció con más crecimiento. Se probaron programas diseñados por expertos, con el docente sostenido todo el curso.

2.1.2. El ensayo que no salió bien, leído entero

Del otro lado está el ensayo de la Education Endowment Foundation, evaluado por la Universidad de Durham (Menzies y otros, 2016). Se asignaron al azar 24 centros ingleses, con 4.074 alumnos de Year 7, el primer curso de la secundaria inglesa (11-12 años). Tras un año, la medida de lengua dio un tamaño del efecto de -0,13, unos dos meses menos de progreso, en conjunto; y de -0,24, unos tres meses menos, en el alumnado con derecho a comedor gratuito.

La conclusión del informe es más prudente que el titular que circula: «Adopting PBL had no clear impact on either literacy [...] or student engagement with school and learning». Sin efecto claro en conjunto, con un posible efecto negativo en el alumnado de familias con menos recursos. Y el propio informe pone tres límites a su resultado. La seguridad del resultado global es baja, y el del subgrupo no tiene calificación de seguridad. Faltan datos del 47 % del alumnado de intervención frente al 16 % del de control, y cinco centros de intervención abandonaron. Además, algunos centros de control hicieron su propia versión de proyectos durante el ensayo, así que la comparación no fue limpia.

El informe añade que el programa se aplicó con fidelidad, pero que exigía un apoyo directivo sustancial y cambios de organización del centro. No prueba que el formato perjudique; tampoco se puede apartar como si no existiera.

2.1.3. Los metaanálisis tampoco coinciden entre sí

Un metaanálisis es otra clase de fuente: no es un experimento, sino una síntesis que combina los resultados de muchos estudios. El de Zhang y Ma (2023) da un efecto global de 0,44, y el de Chen y Yang (2019), de 0,71; con qué base y con qué límites se lee cada cifra es del epígrafe 2.2.

Más importante que la distancia entre las dos cifras es que no coinciden en qué modera el efecto. Qué decidir con esa discrepancia es del epígrafe 2.2. Y la revisión de MDRC calificó la evidencia de prometedora pero no demostrada (Condliffe y otros, 2017).

2.1.4. Por qué todo es compatible

Cuatro ensayos, cuatro medidas distintas

+4 a +7,6 pp Aprobados AP, secundaria superior (EE. UU.) Knowledge in Action. Matriculados o solo presentados	+8 pp Ciencias, 3.º de Primaria (Michigan) ML-PBL. Prueba externa; igual por nivel lector	0,48 y 0,18 Sociales y lectura informativa, 2.º de Primaria Project PLACE. Tamaños del efecto; nada en escritura ni motivación	-0,13 Lengua, Year 7 (Inglaterra) EEF. Sin efecto claro; -0,24 con comedor gratuito
--	--	---	---

Fuente: elaboración propia a partir de Saavedra y otros (2021), informe técnico de ML-PBL (2021) y Krajcik y otros (2023), Duke y otros (2021) y Menzies y otros (2016). pp: puntos porcentuales.

FIGURA 2.1

Las cuatro cifras de la figura no se contradicen, porque no responden a la misma pregunta. Los ensayos de Lucas probaron unidades de currículo concretas, con formación intensa, y midieron el contenido de la materia en la que se trabajó. El de la EEF probó un programa de cambio organizativo del centro y lo midió con una prueba general de lengua, no con el contenido de los proyectos; además perdió casi la mitad de su muestra de intervención y tuvo controles contaminados. No probaron lo mismo ni midieron lo mismo.

Lo que sí sale de leerlos juntos es una conclusión de diseño, y es la del curso: el efecto favorable aparece cuando el proyecto llega con las unidades escritas, los materiales preparados y el docente acompañado. Ninguno de esos programas dejaba al alumnado a solas con una pregunta y un producto. Las lecciones cortas y la instrucción explícita dentro del proyecto, que el epígrafe 2.3 justifica, forman parte de lo que se probó.

ATENCIÓN · EL FOLLETO NO ES EL ESTUDIO

El folleto de Knowledge in Action habla de unos ocho puntos, que valen solo para quien se presentó al examen; sobre los matriculados son cuatro. Cita la cifra del informe o del artículo, y anota sobre qué población se calculó.

2.1.5. Cuatro preguntas antes de citar un estudio

Cuando alguien te ponga un resultado delante, a favor o en contra, pasa siempre las mismas cuatro preguntas. Qué se probó: un programa completo o solo el formato. Frente a qué se comparó. Qué se midió: el contenido trabajado, una prueba general o la opinión del alumnado. Y qué se perdió: centros que abandonaron, datos que faltan, controles que hicieron algo parecido.

Con esas respuestas, la frase defendible es modesta y útil: hay ensayos con resultados claros a favor de programas completos, en etapas y materias concretas; hay uno sin efecto claro y con un posible efecto negativo en el alumnado desfavorecido; y los metaanálisis no coinciden en qué modera el resultado. Que el formato, por sí solo, mejore los resultados no se sostiene. Y hay un dato que no pasa ni la primera pregunta: los porcentajes de retención de la pirámide del aprendizaje. Sus versiones circulan desde hace más de un siglo sin base empírica (Letrud y Hernes, 2018), y las revisiones que han buscado su fuente no la han encontrado (Lalley y Miller, 2007). Su desmontaje, con el de los estilos de aprendizaje y los hemisferios, es del curso 17.

CASO DE AULA · 1.º DE ESO, LENGUA CASTELLANA Y LITERATURA

El departamento propone un proyecto trimestral de revista escolar y lo justifica con «los ensayos de Estados Unidos». La jefa de departamento pasa las cuatro preguntas: esos ensayos midieron ciencias, sociales y exámenes AP, y el único de edad equivalente, el inglés, midió lengua sin efecto claro. No tumba el proyecto: lo reescribe con dos lecciones cortas por semana sobre los saberes que la revista necesita y fija qué criterios se mirarán en cada entrega. El instrumento es cosa del curso 07.

Aplicación práctica

1. **Redacta la ficha de lectura de un ensayo favorable.** Elige Knowledge in Action, ML-PBL o Project PLACE y lee su resumen en ERIC o en la revista: **lo trabajas con material público.** Responde por escrito a las cuatro preguntas del apartado 2.1.5, copia la limitación que el estudio declara con sus palabras y añade dos líneas sobre qué se parece y qué no a tu grupo real. Guárdalo como [ficha-ensayo-favorable.md](#). **Una hora y media.**
2. **Redacta la ficha de lectura del ensayo de la EEF.** Localiza el informe de Menzies y otros (2016) en ERIC y aplícale las mismas cuatro preguntas: **lo trabajas con material público.** Anota las dos cifras de pérdida de muestra, la seguridad de cada resultado y lo que hicieron los centros de control. Termina con qué alumnos de tu grupo se parecen al subgrupo con comedor gratuito y qué harías distinto por ellos. Guárdalo como [ficha-ensayo-eef.md](#). **Una hora y media.**
3. **Escribe tu respuesta a «está demostrado que el ABP mejora los resultados».** Media página, dirigida a un compañero de tu centro: **lo haces solo.** Cita al menos un ensayo a favor y el de la EEF, cada cifra con lo que midió y su limitación, y cierra con lo que sí puedes afirmar sobre el proyecto que piensas montar con tu grupo. Guárdalo como [respuesta-esta-demostrado.md](#). **Una hora.**

2.2. Variables moderadoras: duración, tamaño del equipo, etapa educativa y materia



Qué modera de verdad el resultado: duración, tamaño del equipo, etapa y materia.

En la reunión de departamento alguien trae el dato: «Los proyectos funcionan mejor entre nueve y dieciocho semanas, con equipos de cuatro o cinco». Lo ha leído en un resumen de un metaanálisis y la cifra suena firme. Antes de fijar la duración y cuadrar los equipos por esa regla, pregúntate de dónde sale, sobre cuántos estudios se apoya y qué otro metaanálisis dice lo contrario. La respuesta cambia lo que puedes hacer con esa cifra: un moderador compara estudios entre sí, no prueba nada sobre el tuyo.

2.2.1. Qué es un moderador y qué no puede decirte

Un moderador es una variable que, dentro de un metaanálisis, se asocia a efectos más grandes o más pequeños. El metaanálisis de Zhang y Ma (2023) reúne 66 estudios experimentales o cuasiexperimentales con 190 tamaños del efecto, y obtiene un efecto global de 0,44 frente a lo que llama «enseñanza tradicional», que no define. Es evidencia empírica agregada, y tiene tres límites que hay que leer antes que las cifras.

El primero es la heterogeneidad: un I^2 del 87,4 %, que quiere decir que los estudios dan resultados muy distintos entre sí y que el promedio esconde mucha dispersión. El segundo es el sesgo de publicación: el test de Begg da señal de un posible sesgo, y aunque los autores lo descartan con otro cálculo, la señal está en su propio artículo. El tercero es el que más importa aquí: los moderadores son comparaciones entre estudios, no experimentos. Nadie tomó un mismo proyecto y lo hizo con equipos de cuatro y de siete: se compararon estudios que diferían en eso y en casi todo lo demás. Por eso ningún moderador prueba que cambiar tu proyecto a equipos de cuatro cause el efecto.

2.2.2. Duración: la franja de nueve a dieciocho semanas

Por duración, los proyectos de nueve a dieciocho semanas dan el efecto más alto de la tabla de Zhang y Ma: 0,673, sobre 35 tamaños del efecto. Los de una a ocho semanas se quedan en 0,498; los de una sola sesión, en 0,359; y los de más de dieciocho semanas, en 0,300. Más largo no es, en esta tabla, mejor.

Dos consecuencias. La primera: el proyecto de seis semanas que este curso usa como ejemplo cae en la franja de una a ocho, así que no puede presentarse como la duración óptima. La segunda: la tabla no separa la duración de lo que se hace dentro de ella, empezando por cuántas lecciones cortas y revisiones lleva. La duración total la fijarás en el epígrafe 3.4, con este dato y con el calendario de tu departamento delante.

2.2.3. Tamaño del equipo: el dato más citado y la base más pequeña

El dato de los equipos de cuatro o cinco personas es el más llamativo y el más frágil. Zhang y Ma le asignan el efecto más alto de todos sus moderadores, 0,909, pero sale de 23 tamaños del efecto, porque solo 81 de los 190 informaban del tamaño del equipo. Los equipos de seis o siete, con 42, dan 0,436. Y el grupo de tres o menos da 0,778 con solo dos: nada que se pueda leer.

A esa base pequeña se suma que otro metaanálisis no encuentra el efecto. Chen y Yang (2019) analizan 46 tamaños del efecto de 30 artículos publicados entre 1998 y 2017, con 12.585 alumnos de 189 centros en nueve países, y obtienen un efecto medio de 0,71. Encuentran que el efecto varía con la materia, la ubicación del centro, las horas de instrucción y el apoyo tecnológico, «but not by educational stage and small group size»: ni la etapa ni el tamaño del grupo. Cuando dos metaanálisis discrepan justo en el dato que ibas a usar, lo que tienes no es una regla: es una pregunta abierta. El criterio con que formarás los equipos se trabaja en el epígrafe 3.3.

2.2.4. Etapa y materia, cada una por su lado

Por etapa, la ventaja en Zhang y Ma es para la secundaria superior (0,720). Primaria (0,527) y la secundaria inferior (0,520) quedan prácticamente empatadas, y la universidad se queda muy por debajo (0,116). Son categorías de sistemas educativos de otros países: no las traslades curso por curso al nuestro.

Lo de la materia son, en realidad, dos moderadores distintos. Por área, destacan ingeniería y tecnología (0,619, sobre 62 tamaños del efecto); las ciencias de la vida dan 0,708, pero con solo tres, y humanidades y sociedad se quedan en 0,284. Por tipo de clase, las de laboratorio o experimentales (0,498) superan a las teóricas (0,393). Fundirlos en «el ABP va mejor en materias técnicas y de laboratorio» mezcla dos comparaciones que el artículo hace por separado.

Moderadores de Zhang y Ma (2023), con el tamaño de su base

0,673 Proyectos de 9 a 18 semanas k = 35 · de 1 a 8 semanas: 0,498	0,909 Equipos de 4 o 5 personas k = 23 · solo 81 de 190 informaban del equipo	0,720 Secundaria superior k = 40 · Primaria 0,527 · universidad 0,116	0,619 Área de ingeniería y tecnología k = 62 · humanidades y sociedad: 0,284
---	--	--	---

Fuente: Zhang, L. y Ma, Y. (2023), *Frontiers in Psychology*, 14, 1202728, tabla 4. Tamaños del efecto (SMD) por categoría; k = tamaños del efecto de la categoría, sobre 190. Efecto global 0,44. Son comparaciones entre estudios, no pruebas de causa.

FIGURA 2.2

Léela por la nota, no por la cifra grande: el 0,909 es el valor más llamativo y el que menos base tiene.

Falta la motivación. El metaanálisis de Wijnia y otros (2024) no es de ABP: mezcla aprendizaje basado en problemas, en proyectos y en casos, en su mayor parte de educación superior. Su efecto medio en motivación ($d \approx 0,50$) es menor cuando la medida es de mejor calidad, mayor en una sola asignatura que en un currículo completo, trivial en las razones para estudiar, y llega con indicios de sesgo de publicación. Que el proyecto motive no está garantizado.

ATENCIÓN · LEER LA TABLA DE MODERADORES COMO UNA RECETA

El error es tomar la categoría ganadora de cada fila y sumarlas: doce semanas, equipos de cuatro, Tecnología, laboratorio. Esa combinación no la probó nadie. Cada cifra es un promedio de estudios distintos, con una heterogeneidad muy alta, y otro metaanálisis no encuentra efecto ni de la etapa ni del tamaño del equipo. Úsalas para no elegir a ciegas, nunca para justificar una decisión ante el claustro como si estuviera demostrada.

2.2.5. Infantil y Primaria: cuando la evidencia no permite concluir

Si das clase en Infantil o en Primaria, el dato que más te afecta no es un moderador, sino la revisión sistemática de Ferrero, Vadillo y León (2021) que ya viste en el epígrafe 1.4: once estudios, con datos de 722 alumnos, ninguno con diseño experimental y casi ninguno con constancia de las condiciones en que se trabajó.

Siete de los once informaron de efectos positivos, pero los autores califican los resultados de no concluyentes: la evidencia no permite concluir, ni a favor ni en contra. Para tu aula, las condiciones que casi nadie comprobó —conocimiento previo, trabajo en equipo entrenado, formación docente— son las que tienes que comprobar tú.

2.2.6. De cada moderador, una decisión

Cada moderador se convierte en una decisión del bloque 3. Lo que sigue es criterio de este curso, no hallazgo de ningún estudio:

- **Duración.** Si tu proyecto pasa de un trimestre, justifica por qué; si dura menos de nueve semanas, no lo compenses alargándolo, sino revisando qué lecciones y qué revisiones contiene.
- **Equipo.** No fijas el tamaño por la cifra de Zhang y Ma: fíjalo por lo que tu grupo puede sostener, sabiendo que la evidencia sobre este punto discrepa.
- **Etapa.** En Infantil y Primaria, donde la evidencia no permite concluir, comprueba antes lo que el alumnado ya sabe y programa la guía explícita que verás en el epígrafe 2.3.
- **Materia.** Pregúntate qué parte del proyecto es trabajo de laboratorio o de taller y qué parte es teórica, y dónde necesita cada una una explicación directa.

CASO DE AULA · 4.º DE ESO, TECNOLOGÍA Y DIGITALIZACIÓN

Un departamento decide un proyecto de doce semanas con equipos de cuatro «porque es lo que dice la investigación». La jefa de departamento pide reescribir la justificación. En la nueva versión, las doce semanas se defienden por el calendario y por los hitos, con tres lecciones cortas de explicación directa programadas antes de cada fase de montaje, y los equipos de cuatro, por el número de puestos del taller. Zhang y Ma quedan como orientación, con la nota de que Chen y Yang no encuentran efecto del tamaño del grupo. La decisión es la misma; ahora se sostiene.

Aplicación práctica

1. **Traduce la tabla de moderadores a decisiones para tu proyecto.** Copia las cuatro filas de la figura 2.2 y, al lado de cada una, escribe qué has decidido para tu proyecto y tu grupo real, y por qué: **lo haces solo**. En una tercera columna anota el tamaño de la base de cada dato y si otro metaanálisis lo contradice. Guárdalo como [moderadores-y-decisiones.md](#). **Una hora y media.**
2. **Decide qué haces donde los metaanálisis discrepan.** Lee los resúmenes publicados de Zhang y Ma (2023) y de Chen y Yang (2019): **lo trabajas con material público**. Para la etapa y el tamaño del equipo, escribe qué decisión tomas cuando la evidencia no se pone de acuerdo y con qué criterio propio la sostienes: calendario, espacio, número de alumnos. Guárdalo como [decisiones-donde-discrepan.md](#). **Una hora.**
3. **Escribe la nota de cautela de etapa para tu grupo.** Media página que sitúe tu grupo en la etapa y diga qué evidencia existe para ella y con qué limitaciones; si das clase en Infantil o en Primaria, con los datos de Ferrero, Vadillo y León (2021): **lo haces solo**. Cierra con las comprobaciones previas que harás antes de empezar. Guárdalo como [nota-de-cautela-de-etapa.md](#). **Una hora.**

2.3. Efectos del proyecto sin andamiaje sobre el alumnado con menor nivel previo



Por qué el proyecto sin andamiaje castiga justo al que menos sabe.

Segunda sesión de investigación de un proyecto de 2.º de ESO sobre el agua en tu municipio. Una alumna que ya entiende el ciclo del agua abre tres fuentes, descarta dos en un minuto y empieza a redactar. Su compañero, que no lo entiende, lleva cuarenta minutos saltando de página en página, porque para saber qué buscar tendría que saber ya lo que busca. Los dos han hecho lo mismo: buscar. Solo una ha aprendido algo. La diferencia no está en el esfuerzo, sino en lo que cada uno traía, y de ahí sale el criterio de diseño que usa el resto del curso.

2.3.1. Lo que cuesta el procedimiento, además del contenido

La teoría de la carga cognitiva parte de una asimetría: toda información nueva pasa primero por una memoria de trabajo limitada en capacidad y en duración, y esos límites dejan de pesar cuando la información ya está en la memoria a largo plazo (Sweller, van Merriënboer y Paas, 2019). Es un marco teórico, no un estudio sobre proyectos, y su origen está en el trabajo de Sweller (1988) sobre lo que se separa al experto del novato cuando resuelve problemas.

La teoría distingue dos cargas que en el aula se confunden. La intrínseca depende de la complejidad del contenido. La extrínseca no depende del contenido, sino de cómo se presenta y de «what the learner is required to do by the instructional procedure»: lo que el procedimiento de enseñanza obliga a hacer al alumno. Buscar la fuente, decidir qué parte sirve y organizar lo encontrado es procedimiento. Cuando el alumno todavía no sabe el contenido, esas operaciones consumen la memoria de trabajo que necesitaría para aprenderlo.

Kirschner, Sweller y Clark (2006) lo formulan para la indagación: buscar la información relevante exige mucho a la memoria de trabajo, y mientras se usa para buscar no está disponible para aprender. Trasladarlo al proyecto es una inferencia de este curso, y es directa: el tiempo en que tu alumno busca no es, por sí mismo, tiempo en que aprende.

2.3.2. Por qué la distancia dentro del aula puede crecer

La misma carga no pesa igual en todos. Quien ya tiene el conocimiento previo reconoce la fuente buena y dedica su capacidad a lo nuevo: es la guía «interna» que, según Kirschner, Sweller y Clark, hace retroceder la ventaja de la guía externa. Quien no lo tiene paga el procedimiento entero. Un formato largo y poco guiado reparte el mismo encargo y devuelve resultados distintos, en la dirección de la que partían.

Que el proyecto pueda ampliar esa distancia se sostiene con piezas indirectas. Kirschner, Sweller y Clark resumen un trabajo de Clark (1989) según el cual el alumnado de menor aptitud asignado a tratamientos sin guía puntuó en el postest por debajo del pretest; es una cita de segunda mano. El ensayo de la Education Endowment Foundation dio en el alumnado con derecho a comedor gratuito un tamaño del efecto de $-0,24$ en lengua, unos tres meses menos de progreso, sin calificación de seguridad para ese subgrupo (Menzies y otros, 2016; lo analiza el epígrafe 2.1), y la propia EEF pide prudencia con él, porque no aparecieron efectos negativos en el alumnado de bajo rendimiento en general. Y en un estudio con 53 alumnos de 7.º grado, el de baja capacidad estaba sobrerrepresentado entre quienes bajaban al pasar del grupo a la prueba individual (Webb, 1993, con datos de su informe de 1992; vuelve en el epígrafe 4.2). Ninguna prueba que el proyecto amplíe la distancia. Juntas bastan para no dar por hecho que la reduce.

2.3.3. El argumento de 2006, y lo que no dice

Kirschner, Sweller y Clark (2006) publicaron un artículo de revisión y de posición —no un metaanálisis— contra la instrucción con guía mínima: aquella en que el alumnado, en lugar de recibir la información esencial, tiene que descubrirla o construirla por su cuenta. Meten en ese saco el descubrimiento, el aprendizaje basado en problemas, el experiencial, la indagación y el constructivista. Concluyen que los estudios controlados apoyan casi siempre una guía fuerte con alumnado novato o intermedio, y que su ventaja solo retrocede cuando el conocimiento previo es suficientemente alto.

ATENCIÓN · LO QUE KIRSCHNER, SWELLER Y CLARK NO DIJERON

El artículo no trata el aprendizaje basado en proyectos: la expresión no aparece, y su evidencia sobre el aprendizaje basado en problemas procede sobre todo de Medicina. Citarlo como prueba de que el ABP no funciona es atribuirle lo que no dice. Su argumento afecta a tu proyecto en la medida en que deje al alumnado descubrir o construir por su cuenta la información esencial. Si esa información se enseña, no lo alcanza.

2.3.4. La respuesta, la réplica y lo que comparten

Un año después, en la misma revista, Hmelo-Silver, Duncan y Chinn (2007) respondieron que el aprendizaje basado en problemas y la indagación «are highly scaffolded»: llevan muchas formas de andamiaje y confundirlos con el descubrimiento sin guía es un error. Aportaron estudios de programas completos, con materiales y formación del profesorado, y admitieron que aún no se sabe en qué condiciones rinde más cada entorno andamiado. Schmidt y otros (2007) defendieron, en el mismo número, que el aprendizaje basado en problemas permite adaptar la guía con flexibilidad.

La réplica de Sweller, Kirschner y Clark (2007) concede más de lo que se suele contar: aceptan que el modelado de esos enfoques se parece al ejemplo resuelto y, en esa medida, lo apoyan. Su objeción se queda en la búsqueda innecesaria como método de enseñanza, y en un desacuerdo de método: un ensayo que cambia muchas cosas a la vez no permite saber cuál produjo el efecto.

Lo que comparten los dos bandos es lo que este curso usa: el novato necesita guía explícita dentro del proyecto. Y hay una pieza que admiten ambos: Kirschner, Sweller y Clark citan a la propia Hmelo-Silver (2004), para quien, cuando el alumno necesita un conocimiento concreto en mitad del problema, «a lecture at the right time may be beneficial». La lección corta en el momento justo es el punto en que coinciden quienes defienden el formato y quienes lo critican.

El debate sobre la guía mínima y lo que comparten

¿Cuánta guía necesita quien aprende?

La objeción (2006)

- La guía mínima falla con el novato
- Buscar consume memoria de trabajo
- No trata el ABP

La respuesta (2007)

- Problemas e indagación, muy andamiados
- No son descubrimiento sin guía
- Pruebas: programas completos

La réplica (2007)

- Aceptan el modelado
- Rechazan la búsqueda innecesaria
- Piden cambiar una variable

Lo que comparten

- El novato necesita guía explícita
- Lección en el momento justo
- El modelado vale para ambos

Fuente: elaboración propia a partir de Kirschner, Sweller y Clark (2006); Hmelo-Silver, Duncan y Chinn (2007) y Sweller, Kirschner y Clark (2007).

FIGURA 2.3

2.3.5. El mismo apoyo no sirve a todos: la reversión por pericia

Kalyuga y otros (2003) revisaron el efecto de reversión por pericia: técnicas muy eficaces con alumnado sin experiencia «can lose their effectiveness and even have negative consequences» con alumnado más experto. En el aula: el mismo apoyo que ayuda al novato puede dejar de ayudar, e incluso perjudicar, al que ya sabe. Es una revisión, sin cifra agregada.

El ejemplo resuelto muestra el mecanismo: para el novato rinde más estudiarlo que construir la solución, y a medida que crece la pericia esa ventaja desaparece y después se invierte (Kirschner, Sweller y Clark, 2006). Kalyuga (2007) añade que, con pericia alta, la resolución de problemas poco guiada que consiste en aplicar lo que ya se sabe puede ser el método adecuado. El proyecto poco guiado no está mal en sí: está mal para quien todavía no tiene lo que se le pide aplicar.

2.3.6. Tres consecuencias de diseño

Son inferencias de este curso a partir de lo anterior, no hallazgos de un estudio sobre proyectos.

1. **Enseña la información esencial; no la mandes descubrir.** Lo que hace falta saber para que el proyecto tenga sentido se enseña de forma explícita —lección corta, ejemplo resuelto, modelado—, y la búsqueda se reserva para lo que de verdad hay que aprender a buscar. Dónde van esas lecciones y cómo se retira la ayuda es del epígrafe 2.4.
2. **Recorta la carga del procedimiento para quien parte de menos.** Fuentes seleccionadas en lugar de «buscad en internet», una estructura dada para organizar lo encontrado, decisiones acotadas. No rebaja el contenido: quita búsqueda que no enseña nada.
3. **Gradúa la guía según lo que sabe cada uno, y compruébalo.** El conocimiento previo se mide antes, con la prueba breve del epígrafe 1.4, y la guía se afloja cuando hay pericia, no cuando llega la fecha. Para verlo, cada alumno tiene que dejar evidencia propia, no solo la del equipo: cómo se construye es del epígrafe 4.2, y el instrumento, del curso 07.

Aplicación práctica

1. **Traza el mapa de la carga de tu proyecto.** Escribe, paso a paso, lo que tu alumnado tiene que hacer desde la pregunta guía hasta el producto en el proyecto que estás diseñando: **lo haces solo**. Marca cada paso que obliga a buscar, decidir u organizar, y anota qué conocimiento previo pide y si lo enseñas antes o lo das por supuesto. Guárdalo como [mapa-de-carga-del-proyecto.md](#). **Una hora y media.**
2. **Localiza dónde se atascarían tres alumnos concretos.** Elige a los tres alumnos de tu grupo con menos conocimiento previo sobre el contenido, identificados con un código: **lo haces solo**. Señala sobre el mapa el primer paso en que se quedaría parado cada uno y qué haría mientras tanto su equipo. Guárdalo como [atascos-previstos.md](#). **Una hora.**
3. **Aplica las tres consecuencias de diseño a tu proyecto.** Para cada una, escribe qué cambias: qué información pasas a enseñar y en qué sesión, qué búsqueda recortas y cómo sabrás quién ya no necesita la guía: **lo haces solo**. Contrástalo con los tres atascos anteriores. Guárdalo como [tres-consecuencias-aplicadas.md](#). **Una hora.**

2.4. Enseñanza previa necesaria y retirada progresiva de la ayuda

Qué hay que enseñar antes de mandar investigar, y cuándo se retira la ayuda.

Vuelves al proyecto del agua del epígrafe 2.3, ahora con el consumo de las familias del barrio. Mandas investigar cuánto gasta una familia y cómo podría gastar menos. A los diez minutos, la mitad del grupo copia cifras de la primera web que aparece, tres equipos discuten qué significa «litros por habitante y día» y los dos alumnos que ya dominaban la proporcionalidad han terminado y miran el móvil. Nadie ha hecho nada mal: el diseño mandó descubrir lo que había que explicar. Este epígrafe convierte ese criterio en secuencia: qué se enseña antes, dónde van las lecciones cortas, qué andamios se ponen y cómo se retiran.

2.4.1. El dato que ordena la secuencia, y su alcance

Alfieri y otros (2011) hicieron dos metaanálisis sobre una muestra común de 164 estudios de aprendizaje por descubrimiento. Es evidencia empírica agregada, y dice dos cosas distintas. La primera: cuando se compara el descubrimiento sin ayuda con la instrucción explícita, sale ganando la explícita «under most conditions», en la mayoría de las condiciones ($d = -0,38$; 580 comparaciones). La segunda: el descubrimiento asistido —con retroalimentación, ejemplos resueltos, andamiaje y explicaciones que se le piden al alumno— supera al conjunto de las otras formas de enseñanza con las que se comparó, entre ellas la explícita y el descubrimiento sin ayuda ($d = 0,30$; IC 95 % [0,23; 0,36]; 360 comparaciones).

La segunda cifra es la que se cita mal. No dice que el descubrimiento asistido supere a la instrucción explícita: lo compara con una mezcla que incluye el descubrimiento sin ayuda. Lo que concluyen los autores es que descubrir sin apoyo no beneficia al alumnado y que esas cuatro ayudas sí. Y un límite: es un metaanálisis sobre descubrimiento, no sobre proyectos. Aplicarlo al proyecto es una inferencia de este curso, razonable porque el proyecto mal diseñado deja al alumnado descubrir solo, pero inferencia.

2.4.2. Qué se enseña antes de mandar investigar

El criterio sale del epígrafe 2.3. Kirschner, Sweller y Clark (2006), en una revisión argumentada, llaman guía mínima al entorno en el que el alumno, en lugar de recibir la información esencial, tiene que descubrirla o construirla por su cuenta. De ahí la regla de diseño: **lo que es información esencial para el criterio de evaluación se explica; lo que se investiga es su uso.** En el proyecto del agua, qué es un consumo por persona, cómo se lee una factura y cómo se calcula un porcentaje de reducción se enseñan en lección corta. Qué propuesta de ahorro es viable, eso sí se investiga.

Para un novato, esa explicación tiene respaldo. La misma revisión recoge el efecto del ejemplo resuelto: quien estudia soluciones completas rinde después mejor en problemas nuevos que quien tuvo que resolverlos sin modelo. Rosenshine (2012), en una síntesis divulgativa sin tamaños del efecto, lo traduce al aula: material nuevo en pasos pequeños, con práctica después de cada paso, y modelos antes de pedir que se resuelva solo.

2.4.3. Dónde van las lecciones cortas dentro del proyecto

Las lecciones cortas no van todas al principio, en un bloque que convierte el proyecto en una unidad clásica con epílogo. Van en tres sitios. Antes del lanzamiento, lo imprescindible para entender la pregunta guía. Antes de cada hito que pide un procedimiento nuevo, ese procedimiento. Y a demanda, cuando el atasco es de contenido. Esta última es la que, como viste en el apartado 2.3.4, admiten los dos bandos del debate. Cómo se conduce esa lección a demanda mientras los equipos trabajan es del epígrafe 3.5.

Queda decidir para quién. La prueba de conocimiento previo del epígrafe 1.4 te dice quién es novato en qué, y una lección puede ser para todo el grupo o solo para quien la necesita mientras el resto avanza. Así no castigas al que menos sabe ni haces esperar al que ya sabe.

CASO DE AULA · 2.º DE ESO, MATEMÁTICAS Y GEOGRAFÍA E HISTORIA

Proyecto del agua, seis semanas. La prueba inicial muestra que siete alumnos no calculan porcentajes con soltura. Antes del primer hito, esos siete reciben dos lecciones de quince minutos con una factura resuelta; el resto empieza a recoger datos. Antes del segundo, lección para todo el grupo sobre la gráfica comparativa. En la cuarta semana, un equipo se atasca al pasar de metros cúbicos a litros: diez minutos solo con ese equipo, anotados en el registro.

2.4.4. Qué andamios se ponen

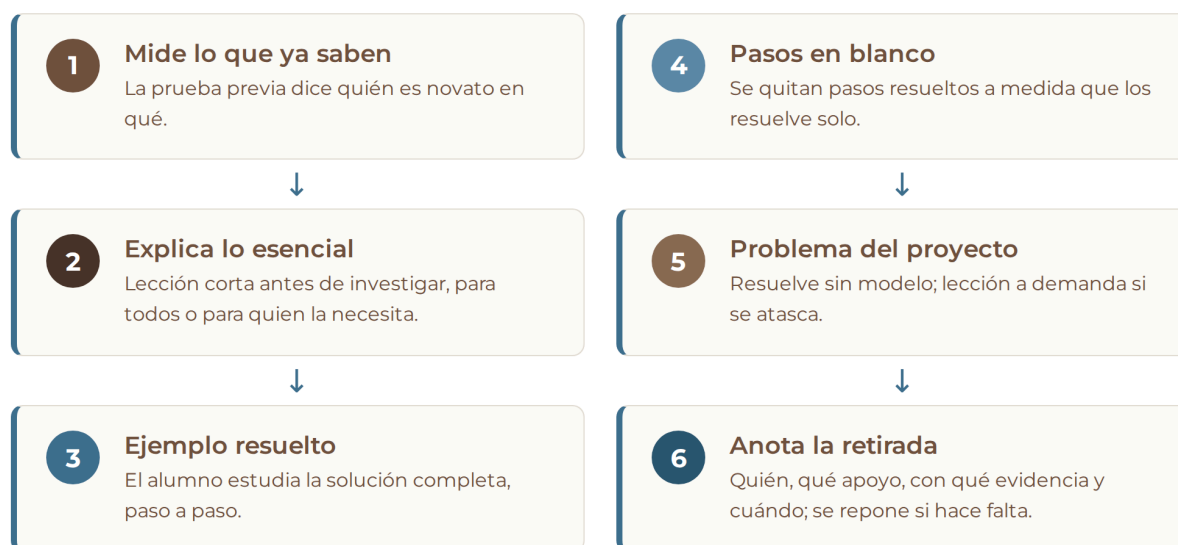
El término viene de Wood, Bruner y Ross (1976), que estudiaron la ayuda de un adulto a un niño en la resolución de un problema, no el trabajo en grupo ni los proyectos: el andamiaje es la ayuda que permite al novato hacer lo que no podría hacer sin ella. Van de Pol, Volman y Beishuizen (2010) revisaron una década de estudios y señalan tres rasgos comunes: contingencia (la ayuda se ajusta al nivel actual del alumno), retirada (disminuye con el tiempo) y transferencia de la responsabilidad. Su conclusión es prudente: los pocos estudios de eficacia sugieren que funciona, y hace falta más investigación.

En un proyecto, los andamios típicos son el ejemplo resuelto, la plantilla, la lista de pasos, la pregunta que orienta la búsqueda y la fuente preseleccionada. El metaanálisis de Belland y otros (2017) da una cifra, con su alcance: el andamiaje por ordenador en STEM, en 144 estudios experimentales desde Primaria hasta adultos, tuvo un efecto positivo consistente ($g = 0,46$). No mide la ayuda que das tú en el aula, y así hay que leerlo.

2.4.5. Cómo se retira, y lo que no sabemos de la retirada

La secuencia de retirada con respaldo más directo es la de Renkl y Atkinson (2003): el ejemplo resuelto es muy eficaz en las primeras fases, y después resolver problemas pasa a ser superior. Proponen, con evidencia empírica a favor, integrar poco a poco pasos de resolución en el ejemplo: del ejemplo resuelto al ejemplo con pasos en blanco, y de ahí al problema. Dejan pendiente cómo ajustar esa retirada al conocimiento previo de cada alumno.

Del ejemplo resuelto al problema, con su registro



Fuente: elaboración propia a partir de Renkl y Atkinson (2003), Kalyuga y otros (2003) y van de Pol, Volman y Beishuizen (2010).

FIGURA 2.4

La figura sitúa esa secuencia dentro del proyecto. La razón para retirar es la reversión por pericia del apartado 2.3.5 (Kalyuga y otros, 2003): el apoyo que ayuda al novato puede dejar de ayudar, e incluso perjudicar, al que ya sabe. Y aquí va la cautela: Belland y otros (2017) no encontraron diferencias según hubiera o no cambio en el andamiaje. La retirada se justifica porque el apoyo que sobra puede estorbar al que ya sabe; que retirar mejora por sí mismo el resultado no está demostrado en general.

ATENCIÓN · RETIRAR A TODO EL GRUPO EL MISMO DÍA

El calendario dice «semana cuatro, sin plantilla», y desaparece para todos. Al que ya resolvía solo le llega tarde; al que aún necesitaba el ejemplo, pronto. La retirada se decide por alumno y con evidencia observable, no por fecha.

2.4.6. Cómo se documenta la retirada

Una retirada que no se anota no se puede revisar ni defender. El registro que propone este curso, como decisión de diseño y no como instrumento validado, tiene cinco columnas: alumno o equipo, andamio, criterio de retirada, fecha y qué pasó. El criterio es observable («resuelve dos ejemplos con pasos en blanco sin pedir ayuda», no «ya lo tiene»), y la última columna admite la marcha atrás: si al quitar el apoyo el alumno se hunde, se repone y se anota.

Ese registro no es una calificación; los instrumentos de evaluación son del curso 07. Es la huella de tus decisiones de enseñanza: te dice en la tercera semana quién sigue con ayuda, y en el epígrafe 4.5 servirá para documentar el proyecto.

Aplicación práctica

1. **Escribe el plan de lecciones cortas de tu proyecto.** Parte de los criterios de evaluación de tu proyecto y de la prueba de conocimiento previo del epígrafe 1.4 aplicada a tu grupo: **lo haces solo**. Separa la información esencial, que se explica, de lo que se investiga. Para cada lección,

anota qué enseña, cuándo va (antes del lanzamiento, antes de qué hito o a demanda), para quién y cuánto dura. Guárdalo como [plan-de-lecciones-cortas.md](#). **Dos horas.**

2. **Diseña la secuencia de andamios con su criterio de retirada.** Elige los dos procedimientos más difíciles de tu proyecto y construye para cada uno el ejemplo resuelto, el ejemplo con pasos en blanco y el problema final, con tu materia y tus datos: **lo haces solo**. Debajo de cada escalón, escribe la evidencia observable que te hará pasar al siguiente y qué harás si el alumno no la da. Guárdalo como [secuencia-de-andamios.md](#). **Una hora y media.**
3. **Prepara el registro de retirada del andamiaje.** Monta la hoja con las cinco columnas del apartado 2.4.6 y rellena ya las filas de tu grupo real con los andamios de la aplicación anterior: **lo haces solo**. Anota al pie cada cuánto la revisarás. Tiene que poder usarse el primer día sin tocarla. Guárdalo como [registro-de-retirada.md](#). **Una hora.**

2.5. Distinción entre trabajo en grupo y aprendizaje cooperativo



Trabajar en grupo y aprender cooperativamente no es lo mismo.

El equipo de cuatro entrega un informe impecable, y tú sabes, porque lo has visto en clase, que lo ha escrito una alumna mientras los otros tres hacían otra cosa. Los cuatro tienen su nombre en la portada y, si no cambias nada, los cuatro tendrán la misma nota. No es mala suerte ni falta de valores del grupo: es lo que produce un diseño en el que el éxito del equipo no depende de lo que aprende cada uno. Sentar juntos a cuatro alumnos no es hacerles cooperar, y la investigación sobre aprendizaje cooperativo pone dos condiciones para que la diferencia se note.

2.5.1. Sentarse juntos no es cooperar

Agrupar es una decisión de organización: quién se sienta con quién y quién firma el producto. Cooperar es una estructura: qué tiene que hacer cada uno para que al grupo le vaya bien, y qué pasa si no lo hace. Un proyecto puede tener equipos durante seis semanas sin tener en ningún momento esa estructura: basta con el formato por defecto, un solo producto firmado por todos.

La literatura sobre el tema es extensa. Johnson y Johnson (2009), en una revisión, hablan de más de 1.200 estudios en once décadas sobre esfuerzos cooperativos, competitivos e individualistas, aunque no todos son estudios de aula. Los mismos autores (Johnson y Johnson, 2008) resumen en cinco los elementos que hacen funcionar la cooperación: interdependencia positiva, responsabilidad individual, interacción cara a cara que ayuda al otro a avanzar, uso adecuado de las habilidades sociales y reflexión del grupo sobre su propio funcionamiento. Sirve para revisar un diseño; las cifras sobre rendimiento vienen de otra síntesis.

2.5.2. Las dos condiciones: objetivos de grupo y responsabilidad individual

Slavin (1996) reunió los estudios sobre métodos cooperativos y rendimiento, y es una revisión con síntesis cuantitativa. De 64 estudios de métodos en los que el reconocimiento al grupo dependía de la suma de lo que había aprendido cada miembro, 50 (el 78 %) encontraron efectos positivos significativos y ninguno los encontró negativos. La mediana del tamaño del efecto fue de +0,32 en los métodos con objetivo de grupo y responsabilidad individual, frente a +0,07 en los que tenían un único producto de grupo o ninguna recompensa de grupo. En los métodos sin objetivos de grupo o con un único producto, solo el 37 % de los estudios halló efectos positivos significativos, y el 14 % los halló negativos.

Tres límites acompañan a estas cifras. Son estudios de 1970 a 1995, sobre todo de Primaria y Secundaria en Estados Unidos, con métodos cooperativos muy estructurados —STAD, TGT, Jigsaw II—, no proyectos: trasladarlas al proyecto es una inferencia de este curso. La «recompensa de grupo» de Slavin es reconocimiento, no la nota, así que **no justifica la nota de grupo**: justifica que el éxito del grupo dependa del aprendizaje de cada uno. Y lo decisivo para ti: el único producto de grupo, que es justo lo que entrega un proyecto si nadie lo remedia, es el formato que peor sale en la síntesis.

Producto único de grupo frente a estructura cooperativa

	Producto único de grupo	Objetivo de grupo y responsabilidad individual
Qué recibe el reconocimiento	El producto que firma todo el equipo	La suma de lo que aprende cada miembro
De qué depende que al grupo le vaya bien	De que el producto salga bien	De que aprendan todos los miembros
Quién responde de cada parte	Nadie en concreto	Cada miembro, ante el grupo
Qué pasa con quien no trabaja	No se ve; otro hace su parte	Se ve y cuenta para el grupo
Estudios con efecto positivo significativo	37 %	78 % (50 de 64)
Mediana del tamaño del efecto	+0,07	+0,32

Fuente: elaboración propia a partir de Slavin (1996), Karau y Williams (1993) y Johnson y Johnson (2008). Cifras de Slavin: estudios de aula de 1970 a 1995 con métodos cooperativos estructurados, no proyectos.

FIGURA 2.5

Entre las dos columnas no cambia el número de alumnos ni el tiempo: cambia de qué depende que al grupo le vaya bien y quién responde de cada parte.

2.5.3. Por qué aparecen el que no hace nada y el que lo hace todo

Slavin lo describe así: en los grupos sin responsabilidad individual, uno o dos alumnos pueden hacer el trabajo del grupo mientras los demás se dejan llevar. Para lo segundo usa un término de la psicología social, la holgazanería social. Lo acuñaron Latané, Williams y Harkins (1979) en dos experimentos en los que las personas aplaudían y gritaban menos fuerte en grupo que solas, una caída del esfuerzo que se sumaba a las pérdidas por mala coordinación. Eran experimentos de laboratorio, con universitarios y tareas físicas.

Karau y Williams (1993) lo definieron como la reducción de la motivación y del esfuerzo cuando se trabaja en colectivo frente a cuando se trabaja solo, y en un metaanálisis de 78 estudios lo encontraron robusto en tareas muy distintas. Entre los moderadores con más peso está la posibilidad de que el esfuerzo de cada uno se evalúe. Es la versión experimental de la responsabilidad individual: cuando lo que hace cada miembro no se ve ni cuenta, baja. El alumno que carga con todo sale del mismo diseño. Ninguno de los dos se arregla con una charla sobre compañerismo: es un problema de diseño.

ATENCIÓN · EL REPARTO DE NOTA QUE PARECE JUSTO

Poner a los cuatro la nota del informe y sumar «un punto por actitud» a quien se ha esforzado. La nota sigue sin decir qué sabe cada uno, y el punto premia lo que se ve, no lo aprendido. La responsabilidad individual no se añade al final: se diseña desde el primer día. Cómo se agrega la evidencia en una calificación es del curso 07.

2.5.4. Repartir en trozos no es el problema

La tentación es prohibir que el equipo se reparta el trabajo. No es la solución. La psicología de grupos distingue entre tareas divisibles, que se pueden partir en subtareas para miembros distintos, y unitarias, que no admiten ese reparto (Steiner, 1972). Es buen vocabulario para describir lo que pides, pero no es un requisito del aprendizaje cooperativo: varios de los métodos que Slavin considera eficaces se basan precisamente en repartir el trabajo, con responsabilidad individual. El reparto en trozos no es malo en sí; lo es cuando nadie responde de lo que ha hecho otro y la nota del grupo no depende de lo que aprende cada uno.

Hay una tensión entre dos autores. Cohen (1994), en una revisión, sostiene que para el aprendizaje rutinario basta con un intercambio limitado de información, pero que en el aprendizaje conceptual con problemas poco estructurados importan la tarea de grupo, la calidad de la interacción y los problemas de estatus dentro del equipo. Slavin pone el acento en la estructura de incentivos; Cohen, en lo que ocurre cuando hablan. Un proyecto tiene tramos de los dos tipos, y lo sensato es no elegir bando. Las partes no delegables de cada miembro y los roles se diseñan en el epígrafe 3.3; las vías de evidencia individual, en el epígrafe 4.2.

2.5.5. Lo que mides decide lo que encuentras

Gijbels y otros (2005) analizaron cuarenta estudios de aprendizaje basado en problemas en educación superior, en aulas reales y frente a la enseñanza convencional, y los ordenaron por el nivel que medía la prueba. Con conceptos sueltos la ventaja fue prácticamente nula (0,068, no significativa); con la comprensión de los principios que enlazan conceptos, clara (0,795, significativa); y en la aplicación a condiciones y procedimientos, positiva pero no significativa (0,339). Las cifras por nivel proceden de la versión de trabajo de los propios autores.

Es aprendizaje basado en problemas y universitario, así que lo que te llevas es la pregunta, no la cifra. Si tu prueba final solo pide definiciones, es muy probable que no registre lo que el proyecto sí haya enseñado; si solo valora el producto, no sabrás qué ha entendido cada uno. Por eso la lección corta sobre los conceptos básicos, antes de que el equipo se ponga a trabajar, no compite con el proyecto: es lo que permite que después haya principios que enlazar. Qué criterio se evidencia en cada momento se decide en el epígrafe 4.1; el instrumento es del curso 07.

Aplicación práctica

- 1. Audita un trabajo en equipo que ya hayas encargado.** Elige un trabajo en grupo de este año o del anterior, con un grupo real: **lo haces solo**. Responde por escrito, sí o no, a cuatro preguntas: si había un objetivo de grupo, si el éxito del grupo dependía de lo que aprendía cada miembro, si alguien respondía de la parte de otro y si el producto era único. Después, identificando a cada alumno con un código, anota en cada equipo quién sospechas que hizo el trabajo y quién no, y qué dato del diseño lo permitió. Guárdalo como **auditoria-tarea-de-equipo.md**. **Una hora y media.**
- 2. Rediseña ese mismo trabajo.** Mantén el producto y cambia la estructura: **lo haces solo**. Escribe qué objetivo de grupo pones, de qué aprendizaje individual depende, qué parte responde cada miembro ante los demás y qué lección corta das antes de repartir. Señala qué decisiones son tuyas de diseño y cuáles se apoyan en Slavin (1996), y deja anotado lo que desarrollarás en los epígrafes 3.3 y 4.2. Guárdalo como **rediseño-tarea-de-equipo.md**. **Una hora.**
- 3. Clasifica lo que mide la prueba de tu proyecto.** Toma la prueba o el guion con el que pensabas comprobar lo aprendido en el proyecto, o la del último que hiciste: **lo haces solo**. Etiqueta cada pregunta como conceptos, principios que enlazan conceptos o aplicación, cuenta cuántas hay de cada nivel y escribe en tres líneas qué nivel queda sin medir y qué consecuencia tiene. Guárdalo como **niveles-de-mi-prueba.md**. **Una hora.**