

# Visibilizando el aprendizaje a través de la autoevaluación

Miguel A. Martínez-Prieto<sup>1</sup>, Jorge Silvestre<sup>1</sup>, Ruth Pinedo-González<sup>2</sup>

<sup>1</sup> Departamento de Informática, Universidad de Valladolid

<sup>2</sup> Departamento de Psicología, Universidad de Valladolid

[miguelamp, jorge.silvestre, ruth.pinedo]@uva.es

## Resumen

Este artículo presenta ADL (*Assessment-Driven Learning*), una dinámica inspirada en la prueba ágil de software que usa autoevaluación y rutinas de pensamiento para retroalimentar y facilitar un análisis continuo del aprendizaje. Los resultados obtenidos muestran que los estudiantes consiguen una buena autopercepción de su aprendizaje, además de activar los movimientos de pensamiento necesarios para razonar sobre él. Asimismo, destacan el impacto positivo de ADL en la calidad de su aprendizaje, a pesar del esfuerzo que les ha supuesto mantener su ritmo de trabajo.

## Abstract

This paper presents ADL (*Assessment-Driven Learning*), an agile testing-based dynamic that uses self-assessment and thinking routines to provide continuous feedback and facilitate critical analysis of learning throughout the subject. Our results show that students achieve a good self-awareness of their learning and activate the necessary thinking moves to reason about it. They also highlight the positive impact of ADL on the quality of their learning, despite the effort required to keep up with their work.

## Palabras clave

Pensamiento visible, autoevaluación, retroalimentación, rutinas de pensamiento, *agile testing*.

## 1. Introducción

“Por desgracia no se puede explicar qué es *Matrix*, has de verla con tus propios ojos [...] Si tomas la pastilla azul, fin de la historia, despertarás en tu cama y crearás lo que quieras creerte. Si tomas la roja, te quedarás en el país de las maravillas [...] Recuerda, lo único que te ofrezco es la verdad, nada más”. Así retaba Morfeo a Neo, en una escena de *Matrix*, a permanecer en la ignorancia o adentrarse en un futuro incierto, pero consciente, en el que vivir la auténtica realidad.

Los estudiantes viven una escena similar al principio de cada asignatura, aunque en su caso sí disponen de

una visión de la auténtica realidad (en términos de objetivos de aprendizaje, contenidos, etc.), que no siempre ayuda a reducir su incertidumbre. Por ello, algunos estudiantes optan por la pastilla azul y transitan por la asignatura cumpliendo, en el mejor caso, con los hitos que establece su sistema de evaluación, pero sin llegar a conseguir un aprendizaje permanente y de calidad.

Esta situación se da en asignaturas en las que los objetivos de aprendizaje y el sistema de evaluación no encajan totalmente. Para evitarlo, el alineamiento constructivo [2] armoniza los objetivos y la evaluación en torno a las acciones formativas, asegurando que dichas acciones movilicen los procesos cognitivos necesarios para alcanzar los objetivos de aprendizaje y, con ello, faciliten que los estudiantes logren un aprendizaje de calidad (y puedan superar el proceso de evaluación establecido). Sin embargo, este alineamiento, por sí solo, tampoco asegura que los estudiantes alcancen la auténtica realidad que promete la pastilla roja, porque pueden seguir sin ser conscientes de su aprendizaje. Según Perkins [9] “el aprendizaje es consecuencia del pensamiento”, por lo que las acciones formativas no solo deben alinearse constructivamente con los objetivos y la evaluación de la asignatura, sino que además deben fomentar que los estudiantes piensen.

En este artículo presentamos una dinámica de aprendizaje guiado por la evaluación (*assessment-driven learning*: ADL), que incorpora la cultura del pensamiento en las actividades de evaluación (pruebas). ADL plantea un proceso continuo de evaluación formativa durante la asignatura, que facilita la retroalimentación del aprendizaje de acuerdo con los resultados obtenidos en las pruebas. Para ello, cada prueba contempla la realización de una tarea alineada con los objetivos de aprendizaje, su autoevaluación y el análisis crítico de sus resultados. Todas las partes de la prueba buscan activar diferentes movimientos de pensamiento [10], enfocados en que los estudiantes obtengan un aprendizaje permanente y de calidad.

El resto del artículo revisa, en la sección 2 los antecedentes de nuestro trabajo y presenta ADL en la sección 3. La sección 4 describe cómo hemos desplegado ADL en una asignatura de Bases de Datos (SBD), cuyos estudiantes han valorado positivamente la metodo-

logía y han obtenido resultados alineados con nuestros objetivos. Finalmente, la sección 5 plantea las conclusiones de esta experiencia.

## 2. Antecedentes

La dinámica de ADL está inspirada en *agile testing* [5], una metodología de prueba basada en los principios ágiles, cuyo objetivo es asegurar la calidad del software producido. *Agile testing* plantea una dinámica de prueba continua y sostenida en el tiempo, que permite identificar errores de forma temprana y retroalimentar constantemente el proceso de desarrollo, habilitando al equipo de trabajo a responder con rapidez ante las incidencias (errores, cambios, etc.) que se produzcan. En *agile testing* se realizan diferentes tipos de pruebas, que se planifican al principio de cada sprint y sirven para asegurar que el software satisface los requisitos del cliente y, en consecuencia, que está listo para su entrega al final del sprint (momento en el que se integra en el proyecto todo el *feedback* generado).

Implantar *agile testing* requiere un entendimiento profundo del producto, lo que demanda equipos con buenas competencias analíticas y de pensamiento crítico [7]. Para abordar estas necesidades, ADL integra el enfoque del pensamiento visible en el marco de la metodología (ágil) de proyectos de aprendizaje.

### 2.1. Proyectos de aprendizaje

Esta metodología de enseñanza-aprendizaje adopta algunas prácticas ágiles, de éxito en los entornos de negocio, para abordar las necesidades existentes en el ámbito educativo. Para ello reinterpreta la asignatura como un proyecto (de aprendizaje) [8] centrado en la construcción del producto de aprendizaje que materializa las competencias que se espera que obtengan los estudiantes.

*Producto de aprendizaje.* Caracteriza el resultado que debe alcanzar un estudiante para culminar el proyecto de aprendizaje. El producto comprende un conjunto de objetivos de aprendizaje que, a su vez, se descomponen en historias de aprendizaje. Cada historia establece un subobjetivo concreto, cuyo aprendizaje aporta valor por sí mismo. La integración de todas las historias pertenecientes al mismo objetivo materializan el aprendizaje esperado en él. Finalmente, los criterios de aceptación establecen el nivel atómico de aprendizaje (son indivisibles) en cada una de las historias. Así, el aprendizaje esperado en cada historia se alcanza al satisfacer todos sus criterios. La especificación de los criterios debe facilitar su medición, ya que son la base sobre la que se evalúa y retroalimenta el producto.

*Proceso de aprendizaje.* El proyecto se divide en sprints de corta duración (2-5 semanas), en los que se

planifican acciones formativas dirigidas a alcanzar el objetivo del sprint. El objetivo del sprint comprende varias historias de aprendizaje (pertenecientes a uno o más objetivos), que empiezan y acaban durante el sprint, y se desarrollan y evalúan de acuerdo con sus criterios de aceptación. Las historias de un mismo objetivo pueden planificarse en sprints diferentes, así que el objetivo finaliza en el último sprint que incluya alguna de sus historias. Los sprints constan de cuatro fases:

- El objetivo del sprint se le comunica a los estudiantes en la fase de “Inicio”, junto con las rúbricas que se utilizarán para evaluar las pruebas.
- Durante la fase de “Trabajo” se ejecutan todas las acciones formativas, incluidas las pruebas.
- En la fase de “Evaluación” se realiza una prueba general sobre el objetivo del sprint, para determinar si este ha sido alcanzado.
- Por último, en la “Retrospectiva” se reflexiona sobre el desarrollo del sprint y se plantean acciones de mejora para el siguiente sprint.

Las acciones formativas y las pruebas se especifican sobre el producto de aprendizaje, asegurando el alineamiento constructivo de la asignatura.

### 2.2. Pensamiento visible

El aprendizaje es un proceso permanente en la vida de las personas y, por tanto, aprender a aprender es una competencia clave para el desarrollo de los estudiantes. Así, nuestra responsabilidad como docentes no es solo que adquieran las competencias propias de la asignatura, sino que también sean conscientes de cómo lo hacen. Siguiendo la metáfora de *Matrix*, tenemos que poner a su disposición la pastilla roja.

Volviendo a la afirmación de Perkins [9], para que los estudiantes aprendan es necesario que lleven a cabo un proceso de pensamiento profundo durante la asignatura. Para alcanzar este objetivo adoptamos el enfoque del pensamiento visible [11], cuyo fin es facilitar que los estudiantes activen diferentes movimientos de pensamiento para obtener un aprendizaje permanente.

*Movimientos del pensamiento.* Son procesos cognitivos que forman parte del pensamiento y la metacognición [10] y cuya activación es vital para lograr una comprensión profunda en el proceso de aprendizaje. El pensamiento visible contempla ocho movimientos: observar de cerca y describir qué hay ahí, construir explicaciones e interpretaciones, establecer conexiones, tener en cuenta diferentes puntos de vista y perspectivas, captar lo esencial y llegar a conclusiones, preguntarse y hacer preguntas, descubrir la complejidad e ir más allá de la superficie, y razonar con evidencia.

*Rutinas de pensamiento.* Son herramientas que ayudan a activar uno o varios movimientos, de acuerdo con un conjunto de preguntas enfocadas en estructurar

el pensamiento de los estudiantes [12]. El éxito de las rutinas radica en su aplicación sistemática durante la asignatura, de forma que los estudiantes adquieran los hábitos de pensamiento que aborda cada una de ellas.

### 3. Aprendizaje guiado por la evaluación (ADL)

La evaluación sumativa se usa para calificar la calidad del producto de aprendizaje, en base a los resultados de las pruebas realizadas y a los criterios establecidos para su valoración. Sin embargo, este tipo de evaluación apenas informa al estudiante sobre las fortalezas o debilidades de su aprendizaje que se esconden tras esa calificación, impidiendo que pueda abordarlas en consecuencia. En contraposición, la evaluación formativa valora el producto durante todo el proceso, generando *feedback* de forma regular. Idealmente, los estudiantes utilizan esa retroalimentación para orientar la construcción de su producto de aprendizaje, de forma que este sea satisfactorio al final del proceso, cuando la evaluación sumativa ya es definitiva. Por lo tanto, la evaluación formativa debe entenderse como un medio didáctico capaz de fortalecer el aprendizaje [4].

ADL propone una dinámica de evaluación formativa continua que se integra en el proyecto de aprendizaje, asegurando la generación de *feedback* durante y al final de cada sprint. Para ello plantea pruebas de diferente alcance, considerando que una evaluación variada ayuda a mejorar los resultados de los estudiantes frente a usar solo exámenes [6]. Las pruebas en ADL comprenden tres partes: la realización de una tarea, su autoevaluación y el análisis crítico de las conclusiones, que se describen a continuación tomando como ejemplo uno de los objetivos de SBD: “Diseño lógico”, que incluye, entre otras, la historia de aprendizaje “Esquema relacional”. Una vez completadas las tres partes, los resultados de la prueba se incorporan en la calificación de los estudiantes.

#### 3.1. Tarea

La tarea es una actividad diseñada para que los estudiantes movilicen los procesos cognitivos que materializan una o varias historias de aprendizaje, dentro del mismo objetivo. Un ejemplo de tarea es resolver un supuesto que solicita obtener el esquema relacional a partir del diseño conceptual de la base de datos, cuyos criterios de aceptación se ilustran en el cuadro 1.

El alcance de las tareas lo determinan los criterios de aceptación de las historias que abordan. Así, ADL plantea tres tipos de pruebas, que se evalúan mediante las rúbricas publicadas en la fase de Inicio del sprint.

*Pruebas básicas.* Abarcan una única historia de

- ☑ LOG-2.1. Transformar cualquier entidad existente en un diseño conceptual (con sus atributos correspondientes) en su representación relacional normalizada (3FN).
- ☑ LOG-2.2. Transformar cualquier relación existente en un diseño conceptual (con sus atributos correspondientes) en su representación relacional normalizada (3FN).
- ☑ LOG-2.3. Establecer el dominio de cualquier columna a partir del diseño conceptual.
- ☑ LOG-2.4. Establecer la clave primaria (PK) de cualquier tabla a partir del diseño conceptual.
- ☑ LOG-2.5. Establecer las claves foráneas (FK) de cualquier tabla a partir del diseño conceptual.
- ☑ LOG-2.6. Nombrar las tablas y columnas a partir del diseño conceptual.

Cuadro 1: Especificación del alcance de la prueba

- [20 %] LOG-2.1. El esquema contiene el 70 % de las tablas necesarias para soportar las entidades consideradas en el diseño conceptual. Cada tabla incorrecta restará en la misma proporción que las tablas correctas.
- [15 %] LOG-2.2. El esquema contiene el 60 % de las tablas necesarias para soportar las relaciones consideradas en el diseño conceptual. Cada tabla incorrecta restará en la misma proporción que las tablas correctas.
- [10 %] LOG-2.3. El esquema declara el 70 % de las columnas necesarias para soportar los atributos considerados en el diseño conceptual. Cada columna incorrecta restará en la misma proporción que las columnas correctas.
- [10 %] LOG-2.4. El esquema establece el 60 % de las PKs necesarias para garantizar la integridad de entidad, de acuerdo con el diseño conceptual. Cada PK incorrecta restará en la misma proporción que las PKs correctas.
- [35 %] LOG-2.5. El esquema establece el 50 % de las FKs necesarias para garantizar la integridad referencial, de acuerdo con el diseño conceptual. Cada FK incorrecta restará en la misma proporción que las FKs correctas.
- [10 %] LOG-2.6. El 60 % de los nombres de las tablas y columnas utilizados en el esquema son consistentes con los usados en el diseño conceptual.

Cuadro 2: Resumen de la rúbrica de evaluación

aprendizaje y se realizan en la fase de Trabajo del sprint, tras finalizar las acciones formativas que materializan la historia. Las pruebas básicas buscan que los estudiantes “conecten” los contenidos presentados con su aplicación para resolver la tarea, y en base a esta experiencia vayan más allá de la superficie, profundicen en la historia de aprendizaje y mejoren su comprensión de la misma. Por ejemplo, una prueba básica en SBD incluye la tarea referida sobre el esquema relacional, cuya rúbrica de evaluación se presenta en el cuadro 2.

*Pruebas de control.* Abarcan parcialmente un objetivo de aprendizaje, por lo que incluyen una o varias de sus historias, pero no todas. Las pruebas de control se realizan en la fase de Evaluación de los sprints en los que se trabaja sobre el objetivo de aprendizaje, excepto aquel en el que finaliza. Las pruebas de control buscan

LOG-2.1	Log-2.2	Log-2.3	Log-2.4	Log-2.5	Log-2.6
Apto	Apto	Apto	Mejorable	No apto	Apto

Cuadro 3: Retroalimentación por criterio de aceptación

activar los mismos movimientos de pensamiento que las básicas, aunque en este caso se realizan tras una mayor exposición a las historias evaluadas. Por ejemplo, una prueba de control en SBD es un examen parcial, en el que se solicita obtener el esquema relacional y asegurar su normalización, en representación de dos de las historias de aprendizaje del objetivo Diseño lógico, cada una con su rúbrica correspondiente.

*Pruebas finales.* Abarcan completamente un objetivo de aprendizaje, por lo que incluyen todas sus historias, y se realizan en la fase de Evaluación del sprint en el que finaliza el objetivo. El proyecto de aprendizaje contempla la realización de una prueba final por cada uno de los objetivos de aprendizaje. En SBD, la prueba final de Diseño lógico añade la historia Fundamentos de diseño lógico a las mencionadas en el ejemplo de las pruebas de control, para así completar el objetivo.

Las pruebas básicas y de control tienen una orientación principalmente formativa, aunque es aconsejable incorporar evaluación sumativa para motivar a los estudiantes a realizarlas. Por otra parte, la evaluación de las pruebas finales debe ser eminentemente sumativa (aunque sin renunciar a la parte formativa), para incorporar sus resultados en la calificación final de la asignatura.

### 3.2. Autoevaluación

La segunda parte de la prueba implica la autoevaluación de la tarea y busca que los estudiantes sean más conscientes del estado de su aprendizaje, considerando que implicarlos en la revisión de su trabajo tiene efectos significativos en su aprendizaje [4]. Cernuda y Riesco [4] presentan un análisis sobre autoevaluación en el que señalan sus limitaciones como medio sumativo, pero enfatizan en su utilidad como medio formativo, dada su capacidad para retroalimentar con prontitud a los estudiantes. En [3] se plantean conclusiones similares y, además, se destaca que las habilidades de los estudiantes para la autoevaluación son limitadas, y se sugiere incorporar rúbricas [1], para centrar su atención en los aspectos importantes.

ADL plantea el uso de dos artefactos para la autoevaluación: la solución y la hoja de corrección.

*Solución.* El proceso de autoevaluación comienza cuando el profesor comparte la solución de referencia de la tarea y les hace llegar a los estudiantes una copia de sus propias soluciones. En este punto, los estudiantes utilizan los criterios establecidos en la rúbrica para comparar ambas soluciones, interpretando si las diferencias entre ellas se deben a que se han seguido

distintas estrategias para resolver la tarea (cuando correspondan) o son errores. Este análisis requiere que el estudiante tenga en cuenta diferentes puntos de vista y se plantee o haga preguntas al respecto.

*Hoja de corrección.* El profesor también comparte una hoja de corrección basada en la rúbrica de la prueba. Así, el estudiante puede registrar los resultados de su autoevaluación y, en base a ellos, obtener *feedback* inmediato, como se ilustra en el cuadro 3. Esta metáfora visual es comparable a la que se usa en la prueba de software, coloreando en verde los criterios de aceptación superados y en rojo aquellos que no alcanzan el nivel de aprendizaje establecido en la rúbrica. Además añadimos el color amarillo con un propósito motivacional, para calificar criterios que alcanzan un nivel de aprendizaje valorable, pero no son aptos. Esta forma de representar los resultados de la prueba busca que los estudiantes no se “conformen” con la nota y usen esta evaluación cualitativa para captar lo esencial de su aprendizaje y llegar a conclusiones. Por ejemplo, del *feedback* mostrado en el cuadro 3 se deduce fácilmente que el estudiante no sabe obtener las claves foráneas (LOG-2.5) y comete errores importantes al establecer las claves primarias (LOG-2.4), habiendo logrado el aprendizaje esperado en el resto de criterios.

Aunque los estudiantes no son conscientes de ello, la hoja de corrección también genera una calificación (sobre 10 puntos) basada en el nivel de aprendizaje alcanzado en los diferentes criterios y en las ponderaciones establecidas en la rúbrica para cada uno de ellos.

### 3.3. Análisis crítico

La tercera parte de la prueba busca que el estudiante razone con las evidencias obtenidas en la autoevaluación y construya explicaciones e interpretaciones que le ayuden a entender el estado de su aprendizaje. Para ello, utilizamos dos rutinas de pensamiento [12].

*Compara-contrasta.* Se usa en las pruebas básicas con el objetivo de que el estudiante analice las similitudes y diferencias encontradas entre su solución y la proporcionada por el profesor. A continuación se le pide que redacte las conclusiones a las que ha llegado y que las concrete sobre los criterios de aceptación, indicando cuáles cree que domina, con independencia del resultado obtenido en la prueba.

*KWL (Know-Want-Learn).* Se usa en las pruebas de control y finales para analizar el aprendizaje obtenido durante el sprint y motivar las expectativas de los estudiantes para el resto de la asignatura. Para ello se pregunta “qué han aprendido durante el sprint” y “cómo lo han hecho” (de acuerdo con las acciones formativas llevadas a cabo), además de “que más quieren aprender” sobre el objetivo de aprendizaje evaluado en la

prueba. Para finalizar, como en la rutina anterior, solicitamos a los estudiantes que indiquen qué criterios de aceptación creen que dominan y, por tanto, qué historias de aprendizaje han alcanzado durante el sprint.

### 3.4. Calificación

La realización de los tres tipos de pruebas asegura el objetivo formativo de ADL, de forma comparable a cómo las pruebas de software ayudan a los desarrolladores en las dinámicas de *agile testing*. A continuación se expone cómo incorporamos el componente sumativo en cada una de las pruebas.

*Pruebas básicas.* El objetivo principal de estas pruebas no es el resultado, sino que los estudiantes las completen en tiempo y forma, para que así puedan identificar problemas en su aprendizaje de forma temprana. Por ello, su calificación se plantea con espíritu motivador y está basada, exclusivamente, en el resultado de la autoevaluación: la nota es la proporcionada por la hoja de corrección, aunque desde las perspectiva de los estudiantes es una calificación asignada por el profesor.

*Pruebas de control.* Estas pruebas actúan a modo de examen parcial de cada objetivo de aprendizaje y su calificación es responsabilidad exclusiva del profesor, que evalúa las soluciones de los estudiantes usando la misma hoja de corrección utilizada en la autoevaluación. Esto le permite obtener una evaluación cualitativa (en los términos ilustrados en el cuadro 3) y una calificación numérica para cada estudiante. En este caso, todos los estudiantes reciben el resultado cualitativo, pero la calificación numérica solo se le entrega a aquellos que hayan alcanzado todos los criterios de aceptación. En caso contrario, la prueba no es apta y su resultado no se tiene en cuenta en términos sumativos.

*Pruebas finales.* El alcance de estas pruebas es el propio del examen final de cada objetivo de aprendizaje y su calificación se realiza de forma comparable al caso anterior. Así, solo los estudiantes que hayan superado todos los criterios de aceptación evaluados reciben la nota que califica su aprendizaje definitivo sobre el objetivo evaluado. Al resto se le notifica que la prueba no ha sido apta y se les comunica que tendrán otra oportunidad en el examen oficial de la asignatura.

## 4. Caso de estudio

Durante el curso 2023-2024 hemos incorporado ADL al proyecto de aprendizaje de una asignatura obligatoria de Bases de Datos (SBD) de 2º curso, cuya carga de trabajo es de 6 ECTS y su producto de aprendizaje contiene tres objetivos principales: Diseño conceptual, Diseño lógico y Manipulación de datos con SQL. El proyecto de aprendizaje de SBD contempla 3 sprints

	SPRINT #1		SPRINT #2		SPRINT #3	
Conceptual	B <sub>1</sub>	C	B <sub>2</sub>	F		
Lógico	B <sub>1</sub>	C	B <sub>2</sub>	F		
SQL			B <sub>1</sub>	C	B <sub>2</sub>	F

Cuadro 4: Distribución de las pruebas realizadas por sprint (*B*: básica, *C*: de control, *F*: final)

de 5 semanas: los dos objetivos de diseño se desarrollan en los sprints #1 y #2, mientras que el objetivo SQL abarca los sprints #2 y #3. Las pruebas vinculadas con cada objetivo se realizan en los sprints en los que están vigentes (ver cuadro 4).

En la presentación de la asignatura planteamos la dinámica de ADL y los diferentes tipos de pruebas, motivando cómo esperábamos que contribuyesen en su proceso de aprendizaje. Como colofón les hablamos de *Matrix* y les ofrecimos las dos pastillas: “la pastilla roja os plantea una dinámica de aprendizaje desconocida, en la que tendréis que evaluar todas las pruebas que realicéis (incluidos los exámenes). Esto os va a suponer un esfuerzo constante durante toda la asignatura pero, a cambio, recibiréis una visión diferente de vuestro aprendizaje. Si los resultados son positivos, podréis consolidar parcial o totalmente vuestro producto de aprendizaje durante el cuatrimestre. En caso contrario, tendréis una nueva oportunidad en el examen oficial, al que podréis presentaros sin penalización alguna”. Por el contrario, “la pastilla azul implica renunciar a participar en las pruebas planificadas durante el proyecto, por lo que podréis seguir la asignatura a vuestro ritmo. Esto no va a penalizaros de forma alguna, ya que podréis realizar el examen oficial de la asignatura en las mismas condiciones que los compañeros que opten por la pastilla roja”.

A continuación se detalla cómo hemos integrado ADL en la asignatura, se presentan algunos resultados de interés y las valoraciones de los estudiantes.

### 4.1. Diseñando las pruebas

Todas las pruebas planteadas este curso incluyeron los mismos tipos de tareas y se distribuyeron temporalmente de forma comparable a los cursos anteriores, de acuerdo con la dinámica de trabajo establecida en el proyecto de aprendizaje.

*Pruebas básicas.* Se implementaron como actividades de resolución de problemas y se programaron justo después de las sesiones de laboratorio en las que se practicaban los conceptos expuestos en las clases de teoría. Por ejemplo, la primera prueba básica sobre Diseño lógico se planificó inmediatamente después de la sesión en la que se enseñó cómo obtener el esquema relacional de una base de datos, a partir de su diseño conceptual. Las pruebas comenzaban los jueves, con la publicación de la tarea, y los estudiantes disponían

	<i>B</i> <sub>1</sub>	<i>C</i>	<i>B</i> <sub>2</sub>	<i>F</i>
Conceptual	76,24 %	85,78 %	86,74 %	92,14 %
Lógico	90,45 %	88,72 %	85,12 %	95,00 %
SQL	92,38 %	76,53 %	90,25 %	85,44 %

Cuadro 5: Autopercepción de los estudiantes

	<i>Control</i>		<i>Final</i>	
	Autop.	Eval.	Autop.	Eval.
Conceptual	85,78 %	76,42 %	92,14 %	86,43 %
Lógico	88,72 %	79,52 %	95,00 %	93,42 %
SQL	76,53 %	55,67 %	85,44 %	76,67 %

Cuadro 6: Autopercepción vs. evaluación

hasta el siguiente martes para resolverlas. Los miércoles se compartían la propuesta de solución, la hoja de corrección y la rutina de pensamiento, y los estudiantes tenían hasta el viernes para hacer la autoevaluación y completar el formulario de análisis. Finalmente, el profesor publicaba los resultados el lunes siguiente.

*Pruebas de control y finales.* Se materializaron en forma de exámenes escritos, en los que se incluyeron problemas comparables a los realizados en las pruebas básicas, aunque con una mayor complejidad a medida que avanzaba el curso. Estas pruebas se realizaron el último día de cada sprint (miércoles) e inmediatamente después de terminarlas el profesor compartía la propuesta de solución, la hoja de corrección y el formulario de la rutina de pensamiento, además de enviar a cada estudiante su examen escaneado. Los estudiantes disponían hasta el siguiente lunes para realizar la autoevaluación y el análisis de resultados, y el profesor publicaba las evaluaciones el martes.

Para implementar ADL hemos diseñado tres hojas de corrección en Excel (una por cada objetivo) y dos formularios en Microsoft Forms (uno para cada rutina).

## 4.2. Conociendo la auténtica realidad

El presente de estudio se realizó con 43 de los 46 matriculados en SBD (3 estudiantes optaron por la pautilla azul), aunque 3 de ellos abandonaron la asignatura al finalizar el Sprint #2 y 2 más no se presentaron a los exámenes oficiales. Por lo que los siguientes datos incluyen los resultados de los 38 estudiantes que completaron la asignatura en la dinámica de ADL.

*Percepción del aprendizaje.* El objetivo principal de ADL es que los estudiantes sean conscientes de la auténtica realidad de su aprendizaje, para conseguir que este sea permanente y de calidad. Un buen estimador de este objetivo es la autopercepción que los estudiantes tienen del aprendizaje, que evaluamos mediante la cuestión “¿qué criterios de aceptación crees que dominan?”, al final de todas las pruebas.

El cuadro 5 muestra el porcentaje promedio de criterios que los estudiantes creen haber superado en cada

prueba. Estos valores son siempre superiores al 75 %, pero concretando en las pruebas finales se observa que una autopercepción va entre el 85,44 % (SQL) y el 95 % (Diseño lógico). Esto significa que los estudiantes creen alcanzar mayoritariamente los criterios, antes de conocer la evaluación del profesor. La comparación de la autopercepción y la evaluación se muestra en el cuadro 6, para las pruebas de control y finales. Aunque los valores de autopercepción son más altos que los de evaluación, la diferencia solo es significativa en la prueba de control de SQL (en el resto las diferencias son inferiores a 10 puntos porcentuales). Por lo tanto, la percepción que los estudiantes tienen de su aprendizaje es muy próxima a la realidad, lo que nos lleva a concluir que han sido capaces de captar lo esencial de su aprendizaje y, con ello, razonar sobre su estado y tomar decisiones para mejorarlo, como se constata en la mejora de la evaluación de la pruebas finales respecto a las de control.

*Calificación.* La calificación de la prueba, que los estudiantes generan inconscientemente durante la autoevaluación, mide la precisión de su corrección respecto a la solución del profesor. Por tanto, es un estimador de su capacidad para analizar los diferentes puntos de vista que materializan ambas soluciones.

La figura 1 muestra las diferencias entre la calificación del profesor y la obtenida por los estudiantes (redondeadas al número entero más cercano): los valores positivos indican que la nota del profesor es superior a la del estudiante y viceversa. En el Diseño conceptual, puede observarse cómo las diferencias se concentran en torno al 0 en la prueba final y apenas el 10 % de las calificaciones difieren en más de 1 punto. Este resultado demuestra que los estudiantes no solo son capaces de crear el diseño conceptual, sino también de interpretar dos diseños conceptuales de una misma base de datos y valorar correctamente sus diferencias. La situación es comparable en el Diseño lógico, aunque aquí las calificaciones de los estudiantes tienden a ser más optimistas en la prueba de control (el 36 % genera una nota superior en 2 o más puntos a la del profesor) y más pesimistas en la final (casi el 20 % genera una nota inferior en 2 o más puntos a la del profesor). Esto se debe a que los estudiantes no penalizan correctamente sus errores en la prueba de control y en la final se van al extremo opuesto.

La situación cambia en el caso de SQL, ya que la concentración en torno al 0 es del 41 % en la prueba de control y apenas de un 36 % en la final. Esto es sintomático de que los estudiantes tienen más dificultades para determinar las equivalencias existentes entre su código SQL y el del profesor. Este resultado no es sorprendente desde la perspectiva de las bases de datos, ya que una misma funcionalidad puede implementarse de diversas formas. Sin embargo, sí puede considerarse un

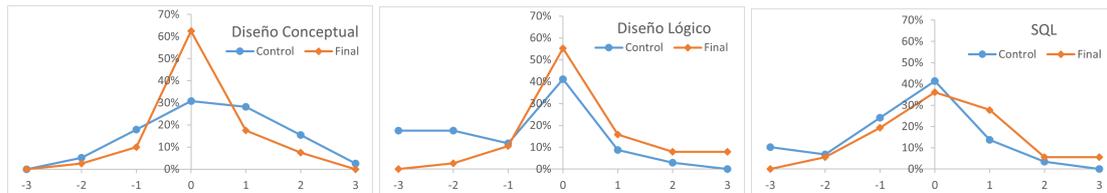


Figura 1: Diferencia entre la calificación (eje X) asignada por el profesor y la obtenida por el estudiante

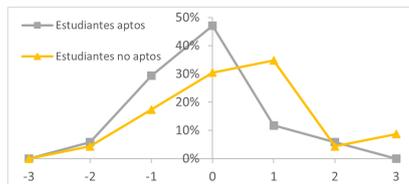


Figura 2: Diferencia de calificaciones (eje X) para estudiantes aptos y no aptos en la prueba final de SQL

déficit desde el punto de vista del aprendizaje, ya que demuestra una falta de comprensión de los operadores de SQL y de cómo usarlos en el ámbito de una consulta. La figura 2 ilustra esta situación con mayor detalle. En ella se observa que la concentración en torno al 0 es más elevada para los estudiantes “aptos”, pero aún así un 35 % de ellos están sobrevalorando su resultados (consideran correctas partes del código SQL que no lo son), mientras que el 78 % de los estudiantes no aptos los infravaloran (no interpretan que algunas partes de su código son equivalentes a las del profesor).

**Resultados académicos.** Solo 4 de los 38 estudiantes participantes en el estudio suspendieron la asignatura, lo que supone una tasa de éxito del 89 %, en línea con los cursos anteriores). Sin embargo, la calificación promedio se ha incrementado hasta el 8,1 (con una desviación típica de 0,61). Solo 1 de los estudiantes que eligió la pastilla azul se presentó a los exámenes oficiales, sin conseguir superar ninguno de ellos.

### 4.3. Valorando la experiencia

Al finalizar el Sprint #3 pasamos una encuesta de valoración que incluía 24 preguntas, principalmente tipo *Likert* (con valoración de 1 a 5), sobre ADL. Obtuvimos 25 respuestas, con las siguientes conclusiones.

**Impacto.** Todos los participantes consideran que la variedad de pruebas que ofrece ADL les ha ayudado a alcanzar los objetivos de aprendizaje, aunque destacan principalmente las básicas (el 92 % le asigna una valoración de 5) frente a las pruebas de control (el 64 % lo valora con 4). Asimismo, destacan que las pruebas les han ayudado a descubrir errores que pasaban desapercibidos (92 % de los participantes) y a comprender conceptos que no tenían claros (80 %). Finalmente, el 88 % de los estudiantes destacan que ADL les ha ayudado “bastante”(4) o “mucho”(5) a alcanzar los ob-

jetivos de aprendizaje de Diseño conceptual y Diseño lógico, reduciéndose este valor hasta el 76 % en SQL.

**Sensaciones.** El 56 % de los estudiantes indica que se sintieron decepcionados o frustrados al ver los resultados de su autoevaluación, pero que estas sensaciones se transformaron en motivación o satisfacción al conocer las valoraciones del profesor (el 88 % se manifiesta de esta forma). Ningún estudiante calificó como “decepcionante” (1) la experiencia vivida al tomar la pastilla roja, mientras que el 84 % indica que la experiencia ha sido “útil” (4) o “satisfactoria (5).

**Esfuerzo.** Los estudiantes consideran de forma generalizada que el esfuerzo extra que ha supuesto tomar la pastilla roja ha merecido la pena, aunque algunos indican que le han dedicado “muchas horas” o que solo la tarea debería ser obligatoria, siendo voluntarias su autoevaluación y análisis crítico. En relación con estas partes de la prueba, el 80 % de los estudiantes indica que le ha resultado “fácil” (4) o “muy fácil” (5) comprender las soluciones de referencia y el 84 % se manifiestan de igual forma respecto a la dificultad de utilizar las hojas de corrección y de interpretar el *feedback* automático que generan.

**Evaluación y feedback.** El 76 % de los estudiantes indica que la evaluación cualitativa proporcionada por ADL les ha permitido conocer de forma precisa el estado de su aprendizaje y el 88 % destaca que ser conscientes de él les ha ayudado a retroalimentar su proceso de aprendizaje de forma “valiosa” (4) o “muy valiosa” (5). Asimismo, el 60 % de los estudiantes indica que la evaluación cuantitativa ahora les parece una medida insuficiente, mientras que el otro 40 % se sigue conformando con saber “si aprueban o suspenden”.

**Movimientos del pensamiento.** Las pruebas de ADL están diseñadas para activar 7 movimientos del pensamiento y, con ellos, visibilizar y fortalecer el aprendizaje de los estudiantes. Como se presenta en la figura 3, el 85 % de los estudiantes señala que ha sido capaz de razonar sobre su aprendizaje en base a las evidencias proporcionadas durante las pruebas, el 77 % valora la contribución que le ha supuesto disponer de diferentes puntos de vista y el 62 % indica que ha captado lo esencial de su aprendizaje y ha llegado a conclusiones al respecto. Estos tres movimientos son fundamentales para generar la consciencia que promete la pastilla roja. Sin embargo, solo el 31 % indica que ha sido capaz

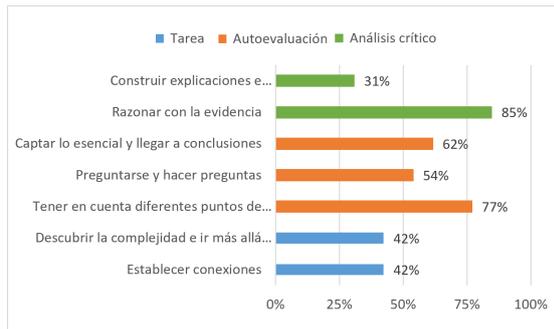


Figura 3: Valoración de los estudiantes sobre los movimientos del pensamiento en ADL

de construir explicaciones e interpretaciones sobre su aprendizaje, lo cual se constata en la longitud y la calidad de las conclusiones que expresan los estudiantes en las rutinas de pensamiento.

## 5. Conclusiones

La dinámica de aprendizaje guiado por la evaluación planteada en ADL ha sido capaz de trasladar al entorno académico las buenas prácticas que sustentan la prueba ágil de software, ofreciendo a los estudiantes *feedback* continuo durante toda la asignatura. Asimismo, el diseño de las pruebas de evaluación sobre el enfoque del pensamiento visible ha permitido a los estudiantes adentrarse en la auténtica realidad de su aprendizaje, ganar consciencia sobre su estado y, con ello, construir productos de aprendizaje de mayor calidad.

Ambos resultados satisfacen los objetivos de nuestro trabajo, aunque los estudiantes manifiestan también que la dinámica de ADL les ha supuesto un esfuerzo importante. Esta percepción está alineada con la sensación de trabajar sin descanso, que alegan los probadores de software [7]. Sin embargo, en nuestro caso solo 9 estudiantes acumularon una carga de trabajo superior a las 150 horas (6 ECTS) contempladas en la asignatura. No obstante, y como parte de nuestro trabajo futuro, revisaremos el alcance y la distribución temporal de las pruebas para mitigar esta sensación. Asimismo, analizaremos la posibilidad de incorporar rutinas de pensamiento específicas para facilitar los movimientos que han reportado unos porcentajes de activación más bajos. Finalmente, también trabajaremos para incorporar herramientas automatizadas que asistan a los estudiantes en la autoevaluación, para así reducir sus imprecisiones y mejorar la comprensión de su aprendizaje.

## Agradecimientos

Este trabajo está vinculado a los PIDs *Aprendizaje guiado por la evaluación en el marco de UVAGILE*

(desarrollado en el marco del GID *Educación ágil y tecnologías aplicadas a la mejora de los procesos de enseñanza-aprendizaje*) y *#PENSA\_TIC Pensamiento crítico y ético a través de las nuevas tecnologías*, subvencionados por el Vicerrectorado de Innovación Docente y Transformación Digital de la Universidad de Valladolid. El segundo autor disfruta de un contrato predoctoral de la Universidad de Valladolid, cofinanciado por el Banco de Santander.

## Referencias

- [1] Heidi Andrade. «Self-assessment through rubrics». En: *Educational leadership* 65.4 (2008), págs. 60-63.
- [2] John Biggs y Catherine Tang. *Teaching for quality learning at university*. 2011.
- [3] Carlos T. Calafate, Juan Carlos Cano y Pietro Manzoni. «Auto-evaluación discente como estrategia de aprendizaje ¿Cuestión de género?» En: *Actas de las JENUI*. 2011, págs. 19-25.
- [4] Agustín Cernuda y Miguel Riesco. «Una Recapitulación sobre la autoevaluación de los alumnos en los estudios de informática: formas, utilidad y aplicación». En: *Actas de las JENUI*. 2011, págs. 81-90.
- [5] Lisa Crispin y Janet Gregory. *Agile testing: A practical guide for testers and agile teams*. 2009.
- [6] Graham Gibbs y Claire Simpson. «Conditions Under Which Assessment Supports Students Learning». En: *Learning and Teaching in Higher Education* 1 (2005), págs. 3-31.
- [7] Kevin Hierro. *Más allá del TDD*. <https://leanmind.es/es/blog/mas-alla-del-tdd/>. 2023.
- [8] Miguel A. Martínez-Prieto, Jorge Silvestre, Aníbal Bregón y Diego García. «Proyectos de Aprendizaje (ágiles)». En: *Innovación docente en Educación Superior*. 2023.
- [9] David Perkins. *La escuela inteligente. Del adiestramiento de la memoria a la educación de la mente*. Gedisa Barcelona, 2018.
- [10] Ron Ritchhart, Mark Church y Karin Morrison. *Hacer visible el pensamiento: Cómo promover el compromiso la comprensión y la autonomía de los estudiantes*. 2014.
- [11] Shari Tishman y Patricia Palmer. «Visible thinking». En: *Leadership compass* 2.4 (2005).
- [12] Shari Tishman, David Perkins y Eileen Jay. *Un aula para pensar: Aprender y enseñar en una cultura de pensamiento*. 1997.