

# La autoevaluación como método de aprendizaje

Daniel Gayo Avello<sup>1</sup> Hortensia Fernández Cuervo<sup>2</sup> Fernando Torre Cervigón<sup>1</sup>

<sup>1</sup> Dpto. de Informática  
Universidad de Oviedo  
C/Calvo Sotelo s/n 33007 Oviedo  
e-mail: {dani,torre}@lsi.uniovi.es

<sup>2</sup> Licenciada en Psicopedagogía  
e-mail: tensi\_f@hotmail.com

## Resumen

Algorítmica y Lenguajes de Programación es una asignatura troncal impartida en la licenciatura de Matemáticas de la Universidad de Oviedo. Aunque algunos alumnos tienen un interés especial por la informática en general y por la programación en particular, muchos estudiantes encuentran la asignatura demasiado compleja. Esta circunstancia es especialmente acentuada en el estudio de la teoría, señalándose estos contenidos como la principal dificultad de la asignatura.

Partiendo de esta situación, los autores decidieron explorar nuevas vías para facilitar a los alumnos el aprendizaje de los contenidos teóricos. Así, se analizó la factibilidad del uso de herramientas de autoevaluación *online* como sistema de autoaprendizaje. Para ello, se seleccionaron dos tipos de herramienta distintos y se llevó a cabo un experimento de investigación educativa a fin de determinar, por un lado, las posibilidades de la autoevaluación como sistema de autoaprendizaje y, por otro, las posibles ventajas de una herramienta sobre la otra.

## 1. Introducción

La asignatura de Algorítmica y Lenguajes de Programación es una asignatura troncal de doce créditos que se imparte en primer curso de la licenciatura de Matemáticas de la Universidad de Oviedo. La mitad de los créditos totales se dedican a contenidos teóricos, repartiéndose los seis restantes entre prácticas de tablero y prácticas de laboratorio.

Aunque algunos de los alumnos matriculados en la asignatura ya han tenido contactos con la informática, para la mayor parte de los estudiantes la asignatura supone su primer acercamiento a los ordenadores, resultándoles relativamente compleja.

En las primeras clases se presentan los conceptos fundamentales (algoritmo, procesador, acción, variable, etc.). Sin embargo, pronto la materia alcanza un mayor grado de dificultad (estructuras de control, subprogramas, recursividad, etc.) y son muchos los alumnos que acusan el cambio.

Conociendo la dificultad de la asignatura para alumnos de primer curso y conscientes de la importancia que la motivación tiene en el proceso de enseñanza-aprendizaje, se desarrolla una página web<sup>1</sup> de la asignatura. A través de dicha web los alumnos pueden acceder a transparencias, apuntes, enunciados y soluciones de ejercicios y exámenes así como a un foro de discusión.

Cuando habían transcurrido unos tres meses del curso se pasó una pequeña encuesta a los alumnos a fin de determinar cómo percibían la asignatura, al profesorado, al grupo clase y a sí mismos. Las principales conclusiones que se extrajeron de este estudio fueron:

1. Los alumnos percibían de forma positiva tanto al profesorado como a la asignatura.
2. Consideraban la asignatura importante e interesante y al mismo tiempo difícil, especialmente la parte teórica.
3. Los alumnos afirmaban estudiar poco o nada, sobre todo la teoría de la asignatura.

Tales resultados causaron cierta perplejidad

---

<sup>1</sup> [www.di.uniovi.es/~dani/asignaturas/index.html#alp](http://www.di.uniovi.es/~dani/asignaturas/index.html#alp)

entre los profesores de la asignatura y les animaron a plantearse formas de abordar el principal problema que traslucía en la encuesta: el escaso estudio de los contenidos teóricos.

## 2. Planteamiento del problema

Simultáneamente al desarrollo de la encuesta anterior se había planteado la posibilidad de ofrecer algún tipo de sistema de autoevaluación mediante pruebas objetivas en la web de la asignatura. De esta manera, los alumnos dispondrían de un instrumento que les permitiría determinar de forma sencilla el grado de aprendizaje que estaban alcanzando.

Este planteamiento, aunque sencillo, planteaba serias dudas. En primer lugar, podía hacer creer, equivocadamente, a los alumnos que la asignatura sería evaluada mediante el uso de tales pruebas objetivas. En segundo lugar, a no ser que las preguntas de tales pruebas fueran modificadas con cierta frecuencia, se podría producir un simple aprendizaje de respuestas en lugar de un aprendizaje significativo de los contenidos. Por último, al tratarse de pruebas cerradas sin posibilidad de explicación alguna sobre la corrección o incorrección de las alternativas podría, en el mejor de los casos, disuadir al alumno de su uso y, en el peor, frustrarle en el estudio de la asignatura.

Por esas razones se decidió buscar opciones alternativas a las pruebas de respuesta múltiple. Tras un período de búsqueda bastante exhaustivo se encontró una herramienta de autoevaluación con un enfoque diferente: *Duck*<sup>2</sup>.

*Duck* es una herramienta desarrollada por el *Biology Computer Resource Center* de la Universidad de Massachusetts Amherst. Básicamente se trata de un conjunto de *scripts* PHP<sup>3</sup> que permiten implementar un sistema de "cursos". Un curso dispone de una o más áreas incluyendo cada área varias preguntas.

Hay tres aspectos fundamentales que diferencian a *Duck* de una herramienta de respuesta múltiple: el tipo de preguntas, la evaluación y la navegación. *Duck* admite preguntas de respuesta múltiple, respuesta corta y

de respuesta extensa (estas últimas pueden ser enviadas al tutor para su revisión). Por otra parte, en *Duck* no se otorga ninguna puntuación a las respuestas puesto que éstas no son ni correctas ni incorrectas; en su lugar, cada respuesta tiene un *feedback* que la matiza y sirve para resolver las posibles dudas del alumno. Por último, en *Duck* no es necesario seguir una navegación lineal, el usuario puede ir a las preguntas que más le interesen, explorar todas las respuestas, retroceder a preguntas anteriores, etc.

Estas características hacían de *Duck* una posibilidad interesante de cara a desarrollar una herramienta de autoevaluación que aportara un valor añadido para los estudiantes.

Así pues, al plantearse la necesidad de explorar nuevas formas de apoyar el aprendizaje de la teoría de la asignatura se vio la posibilidad de emplear herramientas de autoevaluación para tal fin, así como la obligación de analizar tanto herramientas tradicionales de respuesta múltiple como *Duck*. De esta forma sería posible determinar si el uso de tales instrumentos podía favorecer en alguna medida el aprendizaje y, en caso afirmativo, si existía un planteamiento más adecuado que otro.

## 3. Antecedentes

### 3.1 Evaluación y aprendizaje

En la literatura se encuentran múltiples referencias a la influencia de la evaluación en el aprendizaje; sin embargo, aunque todas constatan una mejoría en el aprendizaje de los alumnos, ninguna proporciona pruebas que permitan establecer una vinculación inequívoca entre ambos fenómenos.

Por ejemplo, [1] describe la utilización de una herramienta de respuesta múltiple como sistema de evaluación *online* así como los positivos efectos que tuvo en los resultados de los estudiantes; sin embargo, el objetivo de la aplicación no era el aprendizaje sino la evaluación del alumnado y los autores no hacen ninguna afirmación categórica sobre la relación existente entre evaluación y aprendizaje.

La referencia [2] trata una experiencia similar: el empleo de una herramienta de evaluación por ordenador como sistema de evaluación formativa; nuevamente, no se proporcionan datos que

<sup>2</sup> <http://bcrc.bio.umass.edu/projects/duck/>

<sup>3</sup> <http://www.php.net>

permitan vincular evaluación y aprendizaje.

Por lo que respecta a iniciativas en el ámbito español, [3] presenta algunas conclusiones sobre la utilización de una aplicación informática de autoevaluación con refuerzo; sin embargo, aunque los resultados parecen positivos son puramente cualitativos y, por tanto, difíciles de valorar.

Así pues, el presente artículo pretende aportar algo más de luz a este interesante método de aprendizaje mediante un estudio riguroso que vincule de forma inequívoca la influencia de la (auto)evaluación en el (auto)aprendizaje.

### 3.2. Aprendizaje operante vs. significativo

En este apartado explicaremos brevemente dos tipos de aprendizaje que aparecen implícitos en las herramientas de autoevaluación: el aprendizaje operante [4] y el significativo [5][6].

Podemos decir que aprendizaje operante es aquel en el que la conducta, lo que un individuo hace o dice, se produce en función de sus consecuencias (p.ej. aparcar en lugares permitidos para evitar las multas).

Entendemos, en cambio, por aprendizaje significativo el que se logra al establecerse relaciones entre los conocimientos previos del alumno y los nuevos conocimientos (p.ej. al explicar el cálculo del área de un triángulo a partir de la superficie de un rectángulo).

Al aplicar estos conceptos teóricos al experimento descrito, encontramos que la herramienta tradicional de respuesta múltiple se encuadraría dentro del aprendizaje operante puesto que cada vez que el alumno responde a una pregunta la herramienta le indica si la respuesta es correcta o no provocando el refuerzo o la extinción de la misma. *Duck*, en cambio, se adapta a un aprendizaje significativo puesto que el alumno, al responder, no se encuentra una “evaluación” sino una serie de argumentos sobre la pregunta y la respuesta seleccionada que le permiten elaborar asociaciones que le llevan a cerciorarse sobre la solución más adecuada.

### 4. Formulación de hipótesis

Así, se planteaban dos cuestiones: en primer lugar debía determinarse si el uso de herramientas de autoevaluación *online* mejoraba el aprendizaje de

los alumnos; en segundo, había que comprobar si una aplicación con retroalimentación (aprendizaje significativo) era superior a instrumentos de respuesta múltiple (aprendizaje operante). Así, el experimento debía tratar de demostrar las siguientes hipótesis:

- La utilización de herramientas de autoevaluación en el web mejora el aprendizaje por parte de los alumnos de conceptos teóricos de programación.
- La herramienta de autoevaluación *Duck* resulta más útil para los alumnos que herramientas de respuesta múltiple.

## 5. Metodología

### 5.1. Identificación de las variables

Para el desarrollo del experimento se seleccionaron únicamente dos variables: la herramienta de autoevaluación constituiría la variable independiente mientras que el rendimiento académico en la parte teórica sería la variable dependiente; el principal objetivo de la investigación sería determinar la influencia de la primera sobre la segunda.

### 5.2. Población y muestra

La población objeto del estudio fueron los matriculados en la asignatura de ALP en la Licenciatura de Matemáticas de la Universidad de Oviedo en el curso 2001/2002.

Sobre esta población se llevó a cabo un muestreo aleatorio estratificado<sup>4</sup>, para ello se dividió la población en distintos estratos teniendo en cuenta el sexo<sup>5</sup> de los alumnos y el número de ocasiones en que el alumno estuvo matriculado en la asignatura.

<sup>4</sup> El muestreo aleatorio estratificado implica dividir la población en subgrupos homogéneos y disjuntos para tomar una muestra aleatoria de cada subgrupo.

<sup>5</sup> En varias referencias [7][8][9] se afirma que mujeres y hombres se enfrentan a los ordenadores de forma diferente, no mejor o peor, sino distinta; aun cuando es muy posible que dichas diferencias no se deban tanto al sexo del usuario como a estereotipos sociales asociados a cada género, los autores consideraron una medida preventiva garantizar el mismo número de sujetos de cada sexo en cada grupo.

Así, resultó que la población estaba formada por 23 mujeres y 19 hombres. Además, exceptuando tres personas que se matriculaban por segunda vez en la asignatura, el resto formalizaban matrícula por vez primera. Este hecho también categorizaba a los alumnos por edad, 18 años para los de primera matrícula y 19 para los de segunda.

Una vez determinados los estratos se seleccionaron de forma aleatoria un número fijo de individuos de cada estrato asignándolos a los grupos 1, 2 y 3. La determinación de la naturaleza de cada grupo (control o experimental) también se decidió de forma aleatoria quedando vinculados de esta forma: 1-control, 2-duck y 3-tradicional.

Al término de este proceso se disponía de una muestra de 24 sujetos distribuidos en tres grupos en los que había 4 mujeres y 4 hombres, contando cada grupo con un alumno repetidor.

Durante el desarrollo de la investigación, se produjo muerte experimental<sup>6</sup> en los tres grupos (en ningún caso afectó a los repetidores), finalizando la investigación con 6 sujetos en el grupo de control y 7 en cada grupo experimental.

### 5.3. Diseño del experimento

El experimento es un diseño multivalente pretest-postest en el que hay un grupo de control y un grupo experimental para cada una de las herramientas de autoevaluación (grupos duck y tradicional). Cada grupo consta de 8 individuos cuya asignación ha sido aleatoria. Los tres grupos realizarían un pretest y un postest. El grupo de control no recibiría tratamiento mientras que los grupos experimentales recibirían dos niveles distintos de la variable independiente: el grupo duck utilizaría la herramienta *Duck* y el grupo tradicional la herramienta de respuesta múltiple.

### 5.4. Procedimiento

Las sesiones experimentales tuvieron lugar en una de las salas de ordenadores de la Facultad de Ciencias de la Universidad de Oviedo durante tres viernes consecutivos en horario de mañana.

Durante la primera sesión se pasó un cuestionario a los alumnos para determinar su

<sup>6</sup> La mortalidad experimental es la pérdida de participantes a lo largo de un experimento con grupos.

actitud hacia la asignatura<sup>7</sup>; posteriormente, se les administró un pretest para precisar su rendimiento académico en la parte teórica de ALP con anterioridad al experimento. Dicho pretest consistió en un cuestionario de 20 ítems relativos a los temas “Funciones y subrutinas” y “Recursividad”, cada ítem presentaba 4 posibles respuestas siendo una sola válida. En la evaluación del cuestionario se penalizaba el uso del azar como método de respuesta otorgando 1/2 punto por cada acierto y -1/6 a cada fallo.

A los alumnos se les notificó la forma en que se iba a puntuar dicha “prueba” aunque no se les informó<sup>8</sup> de la naturaleza del experimento; se les hizo creer que participaban en un programa de evaluación de la calidad de la enseñanza a fin de que las respuestas al pretest no se vieran condicionadas por un ambiente de “examen”.

En esta primera sesión, justo a continuación del pretest, los grupos duck y tradicional recibieron su primera sesión de tratamiento; recibiendo la segunda el viernes siguiente.

El tratamiento recibido por el grupo tradicional consistía en la utilización durante una hora de una herramienta similar a una prueba objetiva de respuesta múltiple; dicha herramienta disponía de 40 ítems<sup>9</sup> con 4 respuestas cada uno y

<sup>7</sup> El instrumento utilizado fue una escala de Likert; un método donde las opiniones sobre un concepto varían desde el “total desacuerdo” hasta el “total acuerdo”.

<sup>8</sup> El comportamiento de un individuo, en particular su productividad, puede cambiar considerablemente si es consciente de que se le observa (efecto Hawthorne); este hecho fue detectado por vez primera en una investigación realizada en una fábrica de Hawthorne (Illinois, EEUU) que pretendía determinar el efecto de la iluminación sobre la productividad. La productividad se incrementaba siempre, independientemente del aumento o disminución de la intensidad lumínica; por ello, se llegó a la conclusión de que el hecho de saberse observado era lo que afectaba a la productividad de los trabajadores. Por otro lado, si los individuos saben que pertenecen a un grupo determinado, en especial al de control, pueden producirse efectos de desmoralización (puesto que no se espera mejoría por su parte) o rivalidad (al esforzarse mucho más que de costumbre para “demostrar” al investigador que son “mejores” que los grupos tratados).

<sup>9</sup> Los ítems de las herramientas de autoevaluación, al igual que el pretest y el postest, trataban los temas “Funciones y Subrutinas” y “Recursividad”; un 25% de los ítems de las herramientas aparecían en el pretest y en el postest.

permitía al usuario saber si había “acertado” o “fallado” en su respuesta así como calcular una puntuación total o ver las respuestas correctas.

El grupo duck utilizaba, durante el mismo período de tiempo, una herramienta similar a la del grupo tradicional puesto que disponía de los mismos ítems y respuestas. Sin embargo, presentaba importantes diferencias respecto a la herramienta tradicional; en primer lugar las respuestas no eran correctas o incorrectas sino que su selección proporcionaba una explicación acerca de su validez; además, el alumno no obtenía ningún tipo de puntuación.

Al finalizar la primera sesión se pasó a los alumnos de los grupos duck y tradicional un cuestionario para determinar el grado de satisfacción respecto a su herramienta.

Para terminar el experimento, una semana después de la conclusión del tratamiento se pasó a los alumnos de los tres grupos un cuestionario<sup>10</sup> para determinar el trabajo personal que habían dedicado a la asignatura durante el mes en que se había llevado a cabo la experiencia. Además, se administró a los tres grupos un postest que contenía los mismos ítems que el pretest aunque tanto las preguntas como las respuestas a cada pregunta aparecían en un orden distinto.

En la Tabla 1 se muestra la forma en que se llevaron a cabo las distintas sesiones.

1º Viernes	09.00	Pretest grupos tradicional y duck.
	09.40	1º Tratamiento grupos tradicional y duck.
	11.00	Cuestionario satisfacción grupos tradicional y duck.
	11.15	Salida grupos tradicional y duck. Pretest grupo control.
2º Viernes	09.00	2º Tratamiento grupos tradicional y duck
3º Viernes	10.00	Cuestionario para determinar trabajo personal grupos duck y tradicional.
	10.15	Postest grupos duck y tradicional.
	10.45	Cuestionario para determinar trabajo personal grupo control.
	11.00	Postest grupo control.

Tabla 1. Temporalización de las sesiones

<sup>10</sup> El instrumento utilizado fue una escala de Thurstone.

## 6. Discusión de los resultados

Mediante el experimento se obtuvieron datos para determinar la actitud de los alumnos ante la asignatura, los conocimientos teóricos antes del tratamiento (pretest), los conocimientos teóricos tras el tratamiento o en ausencia de éste (postest), así como el esfuerzo dedicado a la asignatura.

A partir de dichos datos se debía determinar si existían o no diferencias significativas entre las puntuaciones medias obtenidas por los distintos grupos antes y después del tratamiento. Si las diferencias eran estadísticamente significativas y los tres grupos eran aptitudinal y actitudinalmente equivalentes entonces dichas diferencias sólo podrían ser atribuidas al tratamiento. Para el análisis de datos se usó el paquete SPSS.

En primer lugar, hubo que determinar si los grupos eran equivalentes en cuanto a su actitud hacia la asignatura, sus aptitudes (pretest) y el trabajo personal realizado durante las semanas que llevó el experimento. Para comprobar esto se realizó un ANOVA<sup>11</sup> de un solo factor para cada una de las medidas: actitud, aptitud y trabajo personal.

A continuación se debía determinar si la situación del grupo de control en el postest era similar a la mostrada en el pretest, así como comparar los grupos experimentales con dicho grupo de control y con su situación original en búsqueda de mejoras. Todas estas comparaciones se limitan, básicamente, a contrastar las puntuaciones medias obtenidas por los distintos grupos; para ello se emplea la prueba T de Student<sup>12</sup>.

Por último se tenía que comprobar si el grupo

<sup>11</sup> Análisis de la varianza.

<sup>12</sup> La prueba T permite comparar muestras independientes entre sí (p.ej. las notas de dos grupos distintos) o relacionadas (p.ej. las notas de un mismo grupo antes y después de un tratamiento) y determinar si las diferencias entre las medias son significativas o no. Este contraste debe aplicarse de forma unilateral si se pretende determinar una mejoría o superioridad y de forma bilateral si sólo se quieren determinar diferencias entre las muestras. Tanto ANOVA como la prueba T requieren que las muestras sigan una distribución aproximadamente normal, para garantizar este supuesto puede emplearse el test de normalidad de Shapiro-Wilk; asimismo, es necesario que las varianzas de las muestras sean homogéneas, la prueba de Levene suele utilizarse para esta verificación.

duck (que, recordemos, siguió un aprendizaje significativo) había obtenido mejores resultados que el grupo tradicional (aprendizaje operante).

En aras de la brevedad los resultados numéricos se presentan en un apéndice limitándonos aquí a discutir las conclusiones que se pueden extraer de los mismos.

Como se puede apreciar en la primera parte del análisis estadístico, las puntuaciones alcanzadas por los tres grupos en todas las medidas (pretest, posttest, actitud y trabajo) seguían distribuciones normales y presentaban varianzas homogéneas; de esta forma se cumplían los requisitos para aplicar tanto ANOVA como la prueba T de Student.

La aplicación de ANOVA a los resultados obtenidos en el pretest, prueba de actitud y nivel de trabajo mostraron que los tres grupos eran equivalentes en todos los aspectos; es decir, sus aptitudes y actitudes hacia la asignatura eran equivalentes y durante el período de tiempo que duró el experimento los tres grupos estudiaron la asignatura en la misma medida (más bien poco).

Esto quiere decir que las diferencias existentes entre las puntuaciones obtenidas en el posttest no serían achacables ni a diferencias aptitudinales o actitudinales ni a un esfuerzo diferente por parte de los alumnos, tan sólo a la única variable que los diferenciaba: la herramienta de autoevaluación empleada.

La aplicación de la prueba T de Student demostró que, en el posttest, los grupos duck y tradicional presentaron una mejoría significativa respecto al grupo de control y respecto a su situación inicial en el pretest; por el contrario, el grupo de control, aunque empeoró ligeramente no lo hizo de forma significativa (es decir, permaneció sin cambios).

Así pues, queda demostrada nuestra primera hipótesis: la utilización de herramientas de autoevaluación en el web mejora el aprendizaje por parte de los alumnos de conceptos teóricos.

Por lo que respecta a la superioridad de *Duck* frente a herramientas tradicionales, no se demostró, puesto que la diferencia entre los dos grupos experimentales no fue significativa. Sin embargo, esto no quiere decir que ambas herramientas sean equivalentes, sería necesario realizar más experiencias a fin de demostrar o refutar definitivamente la hipótesis.

En las gráficas que se proporcionan también

en el apéndice se pueden apreciar de una forma muy intuitiva los resultados del experimento.

## 7. Conclusiones y trabajo futuro

A la vista de los resultados se puede afirmar, sin lugar a dudas, que la utilización de herramientas de autoevaluación en el Web facilita a los alumnos el aprendizaje de conceptos teóricos de asignaturas de programación.

La cuestión de si es preferible emplear una herramienta de respuesta múltiple o, por el contrario, resulta más adecuada una herramienta que proporcione un *feedback* al alumno aún queda abierta.

Es cierto que el grado de satisfacción fue mayor en el caso del grupo que utilizó el segundo tipo de herramienta, *Duck*, que en aquellos alumnos que emplearon una herramienta tradicional. Sin embargo, los resultados obtenidos en el posttest por ambos grupos fueron estadísticamente equivalentes por lo que no se puede afirmar que haya diferencias significativas en cuanto al aprendizaje adquirido mediante el uso de una u otra herramienta.

No obstante, es preciso señalar que los estudiantes utilizarán estos instrumentos de una forma totalmente voluntaria; por ello, es posible que en una aplicación real sí existan diferencias cuantificables en los resultados obtenidos por los alumnos en función de la herramienta elegida. Por otro lado, habrá que tener en cuenta que iniciativas como la descrita tienden a ser aprovechadas únicamente por los alumnos más destacados [10].

Por tales razones, una vez demostrada la utilidad de las herramientas de autoevaluación como sistemas de autoaprendizaje, los autores pretenden llevar a cabo una nueva investigación que determine en una situación real cuál de los dos enfoques es más adecuado.

## Referencias

- [1] Phil Davies. *Learning Through Assessment: OLAL. On-Line Assessment and Learning*. Conference Proceedings -Flexible Learning-3rd Annual Computer Assisted Assessment Conference, Loughborough, UK. 1999.

- [2] Leith Sly, Léonie J. Rennie. *Computer managed learning: Its use in formative as well as summative assessment*. Conference Proceedings -Flexible Learning- 3rd Annual Computer Assisted Assessment Conference, Loughborough, UK. 1999.
- [3] J.M. Serra, Rafael Serra Simal. *Autoevaluación con refuerzo, como herramienta informática individual de apoyo en el aprendizaje*. III Congreso Edutec. 1997.
- [4] Burrhus F. Skinner. *La conducta de los organismos*. Edición en español de *The Behavior of Organisms: An Experimental Analysis*. 1938/1975.
- [5] David P. Ausubel. *Psicología educativa: un punto de vista cognoscitivo*. Edición en español de *Educational Psychology. A cognitive View*. 1976/1968.
- [6] Mario Carretero. *Constructivismo y educación*. 1993.
- [7] Janet Carter, Tony Jenkins. *Gender and Programming: What's Going On?* Proceedings of the 4th Annual SIGCSE/SIGCUE conference on Innovation and Technology in Computer Science Education. 1999.
- [8] Charles M. Ray, Carolee Sormunen, Thomas M. Harris. *Men's and Women's Attitudes Toward Computer Technology: A Comparison*. Information Technology, Learning and Performance Journal. 1999.
- [9] Leslie J. Francis. *The Relationship Between Computer Related Attitudes and Gender Stereotyping of Computer Use*. Computers & Education, vol. 22(4) 283-289. 1994.
- [10] Curtis A. Carver, Richard Howard. *An Assesment of Networked Multimedia and Hypermedia*. 1995 IEEE/ASEE Frontiers in Education Conference. 1995.

## Apéndice: Análisis estadístico

Nota: Se emplea un nivel de significancia del 95% (0,05)

### Aproximación a la distribución normal de las puntuaciones en el pretest (Shapiro-Wilk)

Grupo	W	Sig.	Conclusión
control	0,938	0,559	0,938 > 1 y 0,559 >> 0,05 → La distribución puede considerarse normal
duck	0,898	0,365	0,898 > 1 y 0,365 >> 0,05 → La distribución puede considerarse normal
tradicional	0,947	0,672	0,947 > 1 y 0,672 >> 0,05 → La distribución puede considerarse normal

### Homogeneidad de las varianzas entre las puntuaciones en el pretest (Levene)

Estadístico de Levene	gl1	gl2	Sig.	Conclusión
0,879	2	19	0,431	0,431 >> 0,05 → Las varianzas son homogéneas

### Equivalencia entre las puntuaciones en el pretest (ANOVA de un solo factor)

	Suma de cuadrados	Grados libertad	Media cuadrática	F	Sig.	Conclusión
intergrupos	17,537	2	8,769	2,645	0,097	0,097 > 0,05 → Aptitudes equivalentes
intragrupos	62,977	19	3,315			
total	80,514	21				

### Aproximación a la distribución normal de las puntuaciones en la escala de Likert para la actitud

Grupo	W	Sig.	Conclusión
control	0,972	0,899	0,972 > 1 y 0,899 >> 0,05 → La distribución puede considerarse normal
duck	0,967	0,857	0,967 > 1 y 0,857 >> 0,05 → La distribución puede considerarse normal
tradicional	0,921	0,471	0,921 > 1 y 0,471 >> 0,05 → La distribución puede considerarse normal

### Homogeneidad de las varianzas entre las puntuaciones en la escala de Likert para la actitud

Estadístico de Levene	gl1	gl2	Sig.	Conclusión
0,391	2	19	0,682	0,682 >> 0,05 → Las varianzas son homogéneas

### Equivalencia entre las puntuaciones en la escala de Likert para la actitud

	Suma de cuadrados	Grados libertad	Media cuadrática	F	Sig.	Conclusión
intergrupos	0,305	2	0,153	0,014	0,986	0,986 > 0,05 → Actitudes equivalentes
intragrupos	200,286	19	10,541			
total	200,591	21				

**Aproximación a la distribución normal de las puntuaciones en la escala de Thurstone para trabajo personal**

Grupo	W	Sig.	Conclusión
control	0,930	0,522	0,930 > 1 y 0,522 >> 0,05 → La distribución puede considerarse normal
duck	0,864	0,214	0,864 > 1 y 0,214 >> 0,05 → La distribución puede considerarse normal
tradicional	0,909	0,416	0,909 > 1 y 0,416 >> 0,05 → La distribución puede considerarse normal

**Homogeneidad de las varianzas entre las puntuaciones en la escala de Thurstone para trabajo personal**

Estadístico de Levene	gl1	gl2	Sig.	Conclusión
1,345	2	17	0,287	0,287 >> 0,05 → Las varianzas son homogéneas

**Equivalencia entre las puntuaciones en la escala de Thurstone para trabajo personal**

	Suma de cuadrados	Grados libertad	Media cuadrática	F	Sig.	Conclusión
intergrupos	8,568	2	4,284	1,013	0,384	0,384 > 0,05 → Trabajo equivalente
intragrupos	71,879	17	4,228			
total	80,448	19				

**Aproximación a la distribución normal de las puntuaciones en el postest**

Grupo	W	Sig.	Conclusión
control	0,912	0,441	0,912 > 1 y 0,441 >> 0,05 → La distribución puede considerarse normal
duck	0,908	0,410	0,908 > 1 y 0,410 >> 0,05 → La distribución puede considerarse normal
tradicional	0,985	0,978	0,985 > 1 y 0,978 >> 0,05 → La distribución puede considerarse normal

**Homogeneidad de las varianzas entre las puntuaciones en el postest**

Estadístico de Levene	gl1	gl2	Sig.	Conclusión
0,341	2	17	0,716	0,716 >> 0,05 → Las varianzas son homogéneas

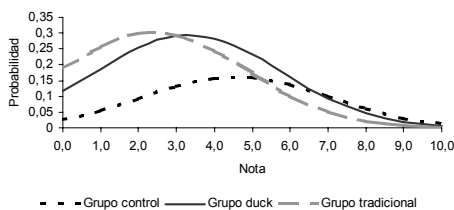
**Mejoría en el postest frente al grupo control y de duck frente al tradicional (prueba T para muestras independientes)**

Grupos comparados	T	Grados libertad	Sig. unilateral	Conclusión
duck vs. control	3,027	11	0,0060	0,0060 << 0,05 → Mejoría significativa
tradicional vs. control	2,965	11	0,0065	0,0065 << 0,05 → Mejoría significativa
duck vs. tradicional	0,265	12	0,796	0,796 >> 0,05 → Mejoría no significativa

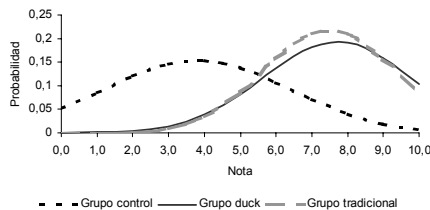
**Cambios respecto a la situación inicial (prueba T para muestras emparejadas)**

Grupos	T	Grados libertad	Sig. bilateral	Conclusión
duck vs. duck	11,664	6	0,000	0,000 << 0,05 → Diferencia significativa
tradicional vs. tradicional	20,457	6	0,000	0,000 << 0,05 → Diferencia significativa
control vs. control	-0,706	5	0,512	0,512 >> 0,05 → Diferencia no significativa

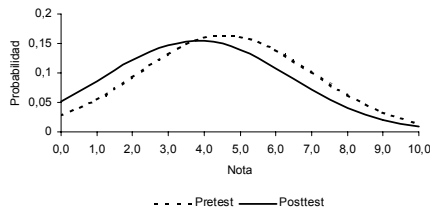
Distribución de las puntuaciones en el pretest



Distribución de las puntuaciones en el postest



Evolución del grupo control



Evolución de los grupos tratados

