



ESCUELA SUPERIOR DE INGENIERÍA

EVALUACIÓN DE COMPETENCIAS  
GENÉRICAS BASADA EN  
INDICADORES PROCEDENTES DE  
REGISTROS DE ACTIVIDADES DE  
APRENDIZAJE

Tesis doctoral presentada por Antonio Balderas Alberico  
dentro del Programa Oficial de Doctorado en Ingeniería y Arquitectura (8104)

Dirigida por Dr. Juan Manuel Dodero Beardo  
y Dr. Manuel Palomo Duarte





ESCUELA SUPERIOR DE INGENIERÍA

EVALUACIÓN DE COMPETENCIAS  
GENÉRICAS BASADA EN  
INDICADORES PROCEDENTES DE  
REGISTROS DE ACTIVIDADES DE  
APRENDIZAJE

Tesis doctoral presentada por Antonio Balderas Alberico  
dentro del Programa Oficial de Doctorado en Ingeniería y Arquitectura (8104)

Dirigida por Dr. Juan Manuel Dodero Beardo  
y Dr. Manuel Palomo Duarte

El doctorando

El director

El director

Puerto Real, Cádiz, julio, 2016



*Evaluación de competencias genéricas basada en indicadores procedentes de registros de actividades de aprendizaje*

Autor: Antonio Balderas Alberico

Director: Juan Manuel Doderó Beardo

Director: Manuel Palomo Duarte

Julio 2016

---



*A María José y Martina.*

Por haberme acompañado en este largo viaje.





## Resumen

El desarrollo de las competencias genéricas es a día de hoy clave en la preparación de estudiantes para su futuro papel en la sociedad. Ante esta situación, los centros de enseñanza han de complementar la docencia del conocimiento específico con el entrenamiento y la evaluación del ejercicio de competencias genéricas. Esto ocurre en un contexto en el que las tecnologías de la información y la comunicación están consolidadas en todos los niveles educativos. Los docentes tienen a su disposición una miríada de entornos virtuales de aprendizaje con los que gestionar y supervisar el proceso de enseñanza de sus estudiantes.

En el análisis de la literatura se han identificado trabajos que utilizan las nuevas tecnologías como apoyo en la evaluación de competencias genéricas. Sin embargo, los métodos mayormente utilizados hasta ahora para la evaluación de dichas competencias presentan los siguientes inconvenientes: problemas de escalabilidad para realizar las evaluaciones con grupos grandes de estudiantes, notables diferencias entre las calificaciones dadas por los estudiantes en procesos de autoevaluación o evaluación entre iguales con respecto a las calificaciones que asignaría el docente, el propósito demasiado específico de algunas herramientas, el esfuerzo necesario para poner en práctica ciertas soluciones, y la aplicabilidad de unos indicadores para medir competencias fuera del contexto en el que se definieron. Por todo esto se detecta la necesidad de desarrollar un método para la evaluación asistida de competencias genéricas.

En esta tesis se propone un método de evaluación denominado *design-based assessment* que propone la utilización de la información de los

registros de actividad de los estudiantes en entornos virtuales de aprendizaje para diseñar indicadores con los que medir su desempeño en competencias genéricas. El método se contextualiza en la metodología de investigación *design-based research*, ya que el diseño de indicadores se lleva a cabo dentro de un proceso iterativo en el que se va refinando el diseño (ciclo de contraste de hipótesis). Para poner en práctica el método se han desarrollado dos lenguajes específicos de dominio con los que facilitar el cálculo de indicadores de dos entornos virtuales de aprendizaje diferentes: sistemas de gestión de aprendizaje y mundos virtuales.

Para la evaluación del método y las herramientas se han realizado seis estudios de caso en tres entornos virtuales de aprendizaje diferentes: wikis, sistemas de gestión de aprendizaje y mundos virtuales. Además, las propuestas realizadas en esta tesis han sido evaluadas por docentes de todos los niveles educativos y pertenecientes a todas las ramas de conocimiento. Los resultados proporcionan, en primer lugar, evidencias a favor de que el método definido permite a los docentes obtener de manera automática valores para un conjunto de indicadores que el diseñador de la evaluación puede expresar a partir de los registros de actividad de los entornos virtuales de aprendizaje. En segundo lugar, también proporcionan evidencias a favor de que las herramientas desarrolladas les permitan tanto diseñar como contrastar estrategias de evaluación a partir de los registros de estos entornos.

## Abstract

The development of generic skills is a key issue for preparing students for their future role in society. Due to this circumstance, educational institutions need to complement the learning of specific knowledge of their subjects with training and evaluation of generic skills performance. This occurs in a context in which information and communications technology are consolidated at all educational levels. Teachers have access to a myriad of virtual learning environments that promote and evaluate the teaching of their students.

In the literature analysis, works have been identified that use new technologies to support the assessment of generic skills. However, the most commonly used methods present the following disadvantages: scalability issues for evaluations when large groups of students are involved, notable differences between the grades given by students in the process of self and peer evaluation when compared to those given by teachers, the high specificity of some tools, the great effort required to implement certain solutions and the applicability of indicators for measuring skills outside the context in which they were created. It is due to these reasons that the development of a method for an assisted evaluation of generic skills is needed.

This dissertation introduces an assessment method called *design based assessment*, which proposes the use of the information registered in the virtual learning environments activity registers to design indicators that measure students' performance in generic skills. The method is contextualized on the research methodology called *design-based research*, as the design of indicators is carried out in an iterative process

in which the design is refined (hypothesis contrast cycle). Two domain-specific languages have been developed to implement this method in two different virtual learning environments: virtual courses and virtual worlds.

Six case studies were conducted for the assessment of the method and tools in three different virtual learning environments: wikis, learning management systems and virtual worlds. In order to contrast the proposals stated in this thesis, they have been assessed by teachers belonging to both all educational levels and all branches of knowledge. From the results, it can be derived that the method allows teachers to automatically obtain a set of indicators of students' activity in virtual learning environments. Additionally, the developed tools allow them to design and contrast evaluation strategies from the registers of these environments.

## **Agradecimientos**

En las siguientes líneas quisiera mostrar mi agradecimiento a todos los que de una u otra manera me habéis ayudado en la culminación de este trabajo:

Gracias María José, por haberte multiplicado por dos en estos últimos meses para que yo pudiera terminar este trabajo y por tu apoyo incondicional.

Gracias mamá, Jose y la chica, por vuestros ánimos y por vuestra ayuda.

Gracias papá, porque aunque hace ya nueve años que no estás, siempre estás presente en todo lo que hago.

Gracia Juanma y Manolo, por haber puesto a mi disposición tantos conocimientos y años de experiencia. Ha sido un placer trabajar con vosotros y espero que vengan muchos proyectos más.

Gracias familia, amigos y compañeros por haber estado ahí y haberos interesado en más de una y dos ocasiones por cómo llevaba la tesis.

Y por supuesto ya estoy disponible para hacer todas aquellas tareas que dejé pendiente para “cuando terminara la tesis”. Gracias por vuestra paciencia.

Antonio

Junio, 2016



# Índice general

<b>Índice de figuras</b>	<b>xiii</b>
<b>Índice de tablas</b>	<b>xvii</b>
<b>Glosario de términos y acrónimos</b>	<b>xxi</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Motivación . . . . .	2
1.2 Contexto . . . . .	6
1.3 Objetivos y preguntas de investigación . . . . .	8
1.4 Estrategia de investigación . . . . .	11
<b>2 Estado del Arte</b>	<b>13</b>
2.1 Preguntas de investigación del SMS . . . . .	14
2.2 Metodología . . . . .	15
2.2.1 Criterios de selección . . . . .	17
2.2.2 Esquema para la extracción de datos . . . . .	18
2.2.3 Visualización y análisis de los datos . . . . .	21
2.3 Resultados . . . . .	21
2.3.1 Localización de la literatura . . . . .	22
2.3.2 Extracción de datos . . . . .	22
2.3.3 Categorización del estudio . . . . .	30
2.3.4 Esquema de clasificación . . . . .	34
2.4 Respuestas a las preguntas de investigación . . . . .	49
2.5 Conclusiones . . . . .	55

## ÍNDICE GENERAL

---

<b>3</b>	<b>Resumen de problemas encontrados</b>	<b>57</b>
3.1	Problemas encontrados . . . . .	57
3.1.1	Evaluación asistida . . . . .	58
3.1.2	Evaluación semiautomática . . . . .	61
3.2	Requisitos . . . . .	62
<b>4</b>	<b>Método para la evaluación de competencias genéricas</b>	<b>65</b>
4.1	Introducción . . . . .	65
4.2	Método: design-based assessment (DBA) . . . . .	66
4.2.1	Contexto . . . . .	66
4.2.2	Descripción del método . . . . .	67
4.3	Características, requisitos y herramienta . . . . .	71
4.3.1	Características . . . . .	71
4.3.2	Requisitos . . . . .	74
4.3.3	Herramienta . . . . .	76
<b>5</b>	<b>Resultados de la investigación</b>	<b>77</b>
5.1	Introducción . . . . .	77
5.2	Antecedentes . . . . .	78
5.2.1	Desarrollo . . . . .	78
5.2.2	AssessMediaWiki (AMW) . . . . .	80
5.3	Lenguajes de dominio . . . . .	90
5.3.1	EvalCourse y SASQL . . . . .	90
5.3.2	EvalSim y VWQL . . . . .	101
5.4	Estudios de caso . . . . .	108
5.4.1	Estudio de caso CS01 . . . . .	108
5.4.2	Estudio de caso CS02 . . . . .	110
5.4.3	Estudio de caso CS03 . . . . .	110
5.4.4	Estudio de caso CS04 . . . . .	110
5.4.5	Estudio de caso CS05 . . . . .	111
5.4.6	Estudio de caso CS06 . . . . .	111
5.5	Cuestionarios de evaluación . . . . .	111
5.5.1	Objetivo . . . . .	112
5.5.2	Distribución . . . . .	113



## ÍNDICE GENERAL

---

5.5.3	Resultados del cuestionario del método y el DSL . . . . .	115
5.5.4	Resultados del cuestionario de indicadores . . . . .	122
5.5.5	Resultados de la encuesta complementaria de indicadores .	128
5.5.6	Discusión y análisis . . . . .	129
<b>6</b>	<b>Conclusiones y trabajo futuro</b>	<b>133</b>
6.1	Conclusiones . . . . .	133
6.1.1	Alcance . . . . .	135
6.1.2	Amenazas a la validez . . . . .	135
6.2	Contribuciones . . . . .	136
6.3	Trabajo futuro . . . . .	140
<b>7</b>	<b>Conclusions and future work</b>	<b>143</b>
7.1	Conclusions . . . . .	143
7.1.1	Scope . . . . .	144
7.1.2	Threats to validity . . . . .	145
7.2	Contributions . . . . .	145
7.3	Future work . . . . .	149
<b>A</b>	<b>Cuestionario de indicadores</b>	<b>153</b>
A.1	Cuestionario . . . . .	154
A.2	Resultados . . . . .	159
A.2.1	Distribución . . . . .	159
A.2.2	Indicadores de participación y de contribución . . . . .	161
A.2.3	Caso de evaluación de competencias genéricas . . . . .	163
A.2.3.1	Conocimientos de programación . . . . .	163
A.2.3.2	Perfil . . . . .	167
A.2.3.3	Comentarios . . . . .	171
<b>B</b>	<b>Cuestionario complementario de la Actuación Avalada</b>	<b>183</b>
B.1	Cuestionario . . . . .	184
B.2	Resultados . . . . .	187

## ÍNDICE GENERAL

---

<b>C Evaluación del método de evaluación (DBA) y del lenguaje para diseñar evaluaciones (SASQL)</b>	<b>193</b>
C.1 Cuestionario . . . . .	194
C.2 Resultados . . . . .	199
C.2.1 Participantes . . . . .	199
C.2.2 Evaluación del método DBA . . . . .	199
C.2.3 Evaluación del DSL . . . . .	208
C.2.4 Evaluación de la sintaxis del DSL (consulta tipo) . . . . .	217
C.2.5 Evaluación de los resultados . . . . .	219
<b>D Listado de competencias genéricas</b>	<b>221</b>
<b>Bibliografía</b>	<b>225</b>

# Índice de figuras

2.1	Distribución de las publicaciones por años . . . . .	24
2.2	Ámbito de trabajos distribuidos según tipo de investigación y según tipo de contribución . . . . .	33
2.3	Competencias genéricas evaluadas con cada método (1 de 2) . . . . .	52
2.4	Competencias genéricas evaluadas con cada método (2 de 2) . . . . .	53
4.1	Diagrama del ciclo de contraste de hipótesis . . . . .	68
4.2	Interacción en el foro en un periodo de tiempo. . . . .	71
5.1	Una rúbrica de ejemplo en AMW . . . . .	80
5.2	Ejemplo de retroalimentación formativa y la contribución de wiki evaluada . . . . .	81
5.3	Ejemplo de flujo de trabajo para la evaluación cualitativa de un wiki utilizando AMW . . . . .	82
5.4	Ciclo de contraste de hipótesis para la evaluación en wikis . . . . .	87
5.5	Metamodelo de SASQL . . . . .	92
5.6	Ciclo de contraste de hipótesis utilizando EvalCourse . . . . .	95
5.7	Interacción en el foro en un período de tiempo. . . . .	96
5.8	Gráfico de contribuciones semanales de cada estudiante a la página del wiki utilizable para evaluar la competencia LEA . . . . .	99
5.9	Ciclo de contraste de hipótesis con EvalSim . . . . .	104
5.10	Distribución de los participantes por rama y sexo . . . . .	116
5.11	Respuestas de los encuestados a la pregunta sobre el método DBA . . . . .	117
5.12	Distribución de los participantes en el cuestionario . . . . .	123

## ÍNDICE DE FIGURAS

---

6.1	Grafo resultante de aplicar técnicas de SNA a un wiki de Moodle . . . . .	141
7.1	Resulting graphic of applying techniques to SNA Moodle wiki . . . . .	150
A.1	Participación de los estudiantes en la página del wiki . . . . .	155
A.2	Contribuciones de los estudiantes a la página del wiki . . . . .	156
A.3	Evolución del contenido de una página del wiki . . . . .	157
A.4	Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el trabajo en equipo . . . . .	164
A.5	Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar la planificación y gestión de tiempo . . . . .	165
A.6	Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el liderazgo . . . . .	166
A.7	Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el trabajo en equipo . . . . .	168
A.8	Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar la planificación y gestión de tiempo . . . . .	169
A.9	Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el liderazgo . . . . .	170
B.1	Docentes que considerarían los indicadores de accesos para evaluar cada competencia genérica de una población de 7 . . . . .	187
B.2	Docentes que considerarían los indicadores de los foros para evaluar cada competencia genérica de un máximo . . . . .	189
B.3	Docentes que considerarían los indicadores de los wikis para evaluar cada competencia genérica de un máximo de 7 . . . . .	190
B.4	Docentes que considerarían los indicadores de las actividades para evaluar cada competencia genérica de un máximo de 7 . . . . .	191
C.1	Resultados consulta . . . . .	197
C.2	Resumen global de respuestas para la característica de la objetividad	202
C.3	Resumen global de respuestas para la característica de la objetividad por rama . . . . .	203
C.4	Resumen global de respuestas para la característica de la adaptabilidad . . . . .	203

## ÍNDICE DE FIGURAS

---

C.5	Resumen global de respuestas para la característica de la adaptabilidad por rama . . . . .	204
C.6	Resumen global de respuestas para la característica de la sistematicidad . . . . .	204
C.7	Resumen global de respuestas para la característica de la sistematicidad por rama . . . . .	205
C.8	Resumen global de respuestas para la característica de la flexibilidad	206
C.9	Resumen global de respuestas para la característica de la flexibilidad por rama . . . . .	207
C.10	Resumen global de respuestas para la característica de la fiabilidad	207
C.11	Resumen global de respuestas para la característica de la fiabilidad por rama . . . . .	208
C.12	Resumen global de respuestas para la característica de la facilidad de aprendizaje . . . . .	211
C.13	Resumen global de respuestas para la característica de la facilidad de aprendizaje por rama . . . . .	211
C.14	Resumen global de respuestas para la característica de la eficacia .	212
C.15	Resumen global de respuestas para la característica de la eficacia por rama . . . . .	213
C.16	Resumen global de respuestas para la característica del ahorro del tiempo . . . . .	213
C.17	Resumen global de respuestas para la característica del ahorro del tiempo . . . . .	214
C.18	Resumen global de respuestas para la característica de la escalabilidad	215
C.19	Resumen global de respuestas para la característica de la escalabilidad por rama . . . . .	215
C.20	Resumen global de respuestas para la característica de la capacidad de reutilización . . . . .	216
C.21	Resumen global de respuestas para la característica de la capacidad de reutilización por rama . . . . .	216
C.22	Resumen global de respuestas para la característica de la fiabilidad	217
C.23	Resumen global de respuestas para la característica de la fiabilidad por rama . . . . .	218



# Índice de tablas

2.1	Resumen de búsqueda de bibliografía . . . . .	17
2.2	Clasificación de trabajos tras aplicar los criterios de selección y exclusión . . . . .	23
2.3	Cantidad de trabajos publicados cada año . . . . .	23
2.4	Competencias genéricas . . . . .	25
2.5	Distribución de publicaciones por tratamiento del problema . . . . .	31
2.6	Competencias genéricas evaluadas mediante autoevaluación y evaluación entre iguales . . . . .	37
2.7	Instrumentos de evaluación y métodos correspondientes a los trabajos de la autoevaluación y evaluación entre iguales . . . . .	38
2.8	Competencias evaluadas directamente por el docente . . . . .	40
2.9	Instrumentos de evaluación y métodos correspondientes a los trabajos en los que la evaluación es realizada por los docentes . . . . .	40
2.10	Competencias evaluadas de forma semiautomática . . . . .	42
2.11	Instrumentos de evaluación y métodos correspondientes a los trabajos de evaluación semiautomática . . . . .	43
2.12	Distribución de publicaciones por tratamiento del problema . . . . .	48
2.13	Número de trabajos que evalúan cada competencia genérica . . . . .	50
2.14	Competencias evaluadas, métodos aplicados y técnicas seguidas para evaluar competencias a partir de los registros de actividad de los entornos de aprendizaje . . . . .	54
4.1	Información sobre la participación en el foro de los estudiantes en un periodo concreto de tiempo . . . . .	70

## ÍNDICE DE TABLAS

---

5.1	Resumen de las competencias evaluadas para cada tipo de indicador	89
5.2	Información sobre la participación en el foro de los estudiantes en un período concreto de tiempo . . . . .	96
5.3	Listado de estudiantes generado por la consulta para la evaluación del CRI . . . . .	100
5.4	Resumen de propuesta de indicadores a partir del uso de EvalCourse	102
5.5	Información sobre las respuestas de una sola palabra dadas por los estudiantes en el role-play . . . . .	105
5.6	Información sobre las palabras por turnos utilizadas por los estudiantes en el role-play . . . . .	106
5.7	Resumen de propuesta de indicadores a partir del uso de EvalSim .	108
5.8	Resumen de competencias medidas en cada estudio de caso . . . .	109
5.9	Resumen de publicaciones de los estudios de caso . . . . .	109
5.10	Resumen de distribución de participantes por rama y sexo . . . . .	116
5.11	Consideración de idoneidad del método DBA por ramas . . . . .	118
5.12	Evaluación de las características del método DBA . . . . .	119
5.13	Consideración de idoneidad del DSL por ramas . . . . .	120
5.14	Evaluación de las características del DSL . . . . .	121
5.15	Consideración de la consulta como entendible o no por ramas . . .	122
5.16	Consideración de los resultados como entendibles o no por ramas .	123
5.17	Perfil de los participantes que participaron en el cuestionario . . .	124
5.18	Resumen de la validez de cada indicador según los participantes en el cuestionario para evaluar cada competencia genérica . . . . .	125
5.19	Cuadro de indicadores considerados para cada competencia . . . .	128
A.1	Resumen de participantes en el cuestionario . . . . .	161
A.2	Perfil de los participantes en el cuestionario . . . . .	162
A.3	Respuestas dadas a la consideración como indicadores de actividad	162
A.4	Resumen de la validez dada por los participantes a los indicadores para evaluar competencias genéricas . . . . .	163
C.1	Selección de características del método DBA . . . . .	195
C.2	Selección de características del lenguaje SASQL . . . . .	196
C.3	Resumen de distribución de participantes por rama y sexo . . . . .	199



## ÍNDICE DE TABLAS

---

C.4	Consideración de idoneidad del método DBA por ramas . . . . .	200
C.5	Evaluación de las características del método DBA . . . . .	201
C.6	Consideración de idoneidad del DSL por ramas . . . . .	208
C.7	Evaluación de las características del DSL . . . . .	210
C.8	Consideración de la consulta como entendible o no por ramas . . .	217
C.9	Consideración de los resultados como entendibles o no por ramas .	219
D.1	Competencias genéricas . . . . .	223



# Glosario de términos y acrónimos

**AMW** AssessMediaWiki

**CEFR** Marco común de referencia europe (Common European Framework of Reference)

**CTMTC** Comprehensive Training Model of the Teamwork Competence

**DBA** Evaluación basada en el diseño (Design-Based Assessment)

**DBR** Investigación basada en el diseño (Design-Based Research)

**DSL** Lenguaje específico de dominio (Domain-Specific Language)

**e-portfolio** Portfolio digital (electronic portfolio)

**GBL** Aprendizaje basado en juegos (Game-Based Learning)

**LMS** Sistema de gestión de aprendizaje (Learning Management System)

**MDE** Ingeniería dirigida por modelos (Model-Driven Engineering)

**MOOC** Cursos online masivos y abiertos (Massive Open Online Courses)

**PBL** Aprendizaje basado en problemas (Problem-Based Learning)

**SASQL** Simple Assessment Specific Query Language

**SMS** Estudio de mapeo sistemático (Systematic Mapping Study)

## **GLOSARIO DE TÉRMINOS Y ACRÓNIMOS**

---

**SNA** Análisis de redes sociales (Social Network Analysis)

**SQL** Structured Query Language

**TBL** Aprendizaje basado en equipos (Team-Based Learning)

**TEL** Aprendizaje mejorado por la tecnología (Technology Enhanced Learning)

**TIC** Tecnologías de la Información y la Comunicación

**VLE** Entornos de aprendizaje virtual (Virtual Learning Environment)

**VWQL** Virtual World Query Language

# Introducción

Las empresas requieren que la formación adquirida por sus nuevos empleados se haya enfocado en competencias genéricas que complementen su formación profesional. Numerosos estudios afirman que el entrenamiento de competencias genéricas como el trabajo en equipo, el liderazgo o las habilidades interpersonales entre otras, tienen un efecto positivo en la adquisición de conocimiento profesional y en la empleabilidad de los estudiantes [40, 71]. La investigación educativa muestra que dichas competencias se enseñan con mayor eficacia en el periodo formativo del individuo [33].

La evaluación de competencias genéricas se ha abordado desde diferentes perspectivas. En la revisión de la literatura realizada por Curtis [32] se identificaron cuatro enfoques diferentes: el primero era la evaluación desde el punto de vista del profesorado, un enfoque muy eficaz sobre todo a nivel escolar, donde el docente conoce de primera mano las características de sus estudiantes, pero que resulta difícilmente transferible; el segundo enfoque consiste en la evaluación mediante portfolios de los estudiantes, enfoque en el que se hace plenamente consciente al estudiante del desarrollo de sus habilidades y que ofrece un relato muy detallado de sus logros, pero que no está en un formato que sea fácilmente digerible ni comparable para el docente; el tercer enfoque es la evaluación basada en la experiencia laboral, que parece ser un método útil para producir una evaluación rápida, pero

## 1. INTRODUCCIÓN

---

al igual que ocurre con la evaluación mediante portfolio, es difícilmente estandarizable y comparable; y por último la evaluación mediante el uso de instrumentos de evaluación estándares, que aunque proporciona una evaluación eficiente y fácilmente interpretable tanto por los estudiantes como por los empleadores potenciales, presenta como desventajas que desacopla la evaluación de la enseñanza siendo un aprendizaje más sumativo que formativo.

Aunque esta revisión de la literatura es relativamente antigua, pues data del año 2004, los métodos actuales no son muy diferentes. La diferencia más importante y que marca el devenir actual en la educación es el protagonismo que adquieren las *Tecnologías de la Información y la Comunicación* (TIC), y que trae como consecuencia que el contexto, los métodos de enseñanza y evaluación se hayan tenido que adaptar. En una época en la que los adolescentes son nativos digitales, incorporar la tecnología a la educación aporta una serie de beneficios que ayudan a mejorar la eficiencia y la productividad en el aula, así como a aumentar el interés de los estudiantes en las actividades académicas [6]. Además, los docentes pueden beneficiarse mucho de los avances tecnológicos para hacer su trabajo más atractivo y ser más eficientes en todas sus actividades cotidianas, entre ellas la evaluación.

Esta tesis surge con la idea de abordar la evaluación de competencias genéricas mediante el uso de herramientas informáticas con un enfoque que se irá desgranando en los próximos capítulos y que se basa en indicadores procedentes de los registros de dichas herramientas.

### 1.1 Motivación

#### **Las competencias genéricas en el marco actual**

En la actualidad, para que la empleabilidad de los nuevos egresados satisfaga las necesidades del mercado laboral europeo, las competencias juegan un papel fundamental [29], y tanto las aptitudes y como las habilidades que la sociedad demandará a los futuros profesionales constituyen pilares básicos a tener en cuenta en el diseño de las estrategias educativas [34]. Por consiguiente, tanto en la enseñanza como la evaluación de competencias deben tenerse en cuenta en los planes de estudio a todos los niveles educativos, incluida la universidad. En el contexto del

Espacio Europeo de Educación Superior <sup>1</sup> e influenciado por la situación actual de la sociedad, sus instituciones sociales y políticas, la universidad se encuentra en el foco de las reformas para alcanzar la convergencia a nivel europeo. En este marco, son las competencias, las tareas y su evaluación los pilares en los que se basa el nuevo currículum universitario [106].

Centrándonos en la evaluación, podemos decir que el foco de interés se centra ahora en cómo evaluar a los estudiantes en el desempeño de sus competencias. Proyectos como el *Tuning Educational Structures in Europe* [52], apoyado por el Lifelong Learning Programme de la Unión Europea [41], muestran la importancia de utilizar el concepto de competencia como base para los resultados de aprendizaje. Las competencias son habilidades que un alumno ha de ser capaz de demostrar una vez que termina su formación. Estas competencias de aprendizaje se dividen en dos grupos: específicas y genéricas [96]. Competencias específicas son aquellas relacionadas directamente con la utilización de conceptos, teorías o habilidades propias de un área en concreto, mientras que las competencias genéricas son habilidades, capacidades y conocimientos que cualquier estudiante debería desarrollar independientemente de su área de estudio [52].

### Las TIC en la educación

En los últimos años, han sido numerosos los avances en lo que al uso de las TIC se refiere. Esto, junto con el asentamiento de Internet, ha traído consigo que la sociedad se haya visto obligada a abordar cambios en su habitual *modus operandi*. Desde la manera en que los ciudadanos interactúan con las instituciones públicas hasta la forma en que estos se relacionan con sus amigos. Y por supuesto, también ha afectado a la educación. El campo de investigación que aborda el uso de la tecnología como parte del proceso de aprendizaje es el *aprendizaje mejorado por la tecnología* (TEL, del inglés *Technology Enhanced Learning*).

Según la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), es preciso que los docentes reciban los instrumentos necesarios para alcanzar los objetivos sociales y económicos que constituyen el eje

---

<sup>1</sup><http://www.eees.es/>

## 1. INTRODUCCIÓN

---

de todo sistema educativo nacional. Para ello, se ha definido el marco de competencias de los docentes en materia de TIC de la UNESCO [72], un conjunto de directrices que definen las competencias necesarias para impartir una enseñanza eficaz mediante el uso de las TIC. Este marco tiene por objeto informar de la función de las TIC en la reforma educativa, así como ayudar a los Estados Miembros a que elaboren criterios de competencia en la materia para los docentes. Herramientas como los sistemas de gestión de aprendizaje (LMS, del inglés *Learning Management System*), los wikis o los mundos virtuales son más que habituales como soporte a la docencia presencial (*blended learning*), y en algunos casos, como ocurre con los cursos online masivos y abiertos (MOOC, del inglés *Massive Open Online Courses*), son el único punto de encuentro entre el estudiante y el docente.

En la mayoría de estas herramientas la actividad generada por cada estudiante suele quedar registrada. Entiéndase por *actividad generada* a la información de la interacción que cada estudiante realiza con la herramienta como, por ejemplo, los accesos al sistema, el envío de sus actividades o la lectura de un mensaje en un foro. Según [25, 47] la recopilación de los rastros de interacción producidos por este tipo de herramientas, con un filtrado adecuado, podría ser una información muy valiosa para obtener indicadores del desempeño de los estudiantes en ciertas competencias. La explotación de este tipo de información puede ser llevada a cabo mediante técnicas de *Learning Analytics* [30]. El *learning analytics* es un área dentro del TEL que está enfocada en el desarrollo de métodos para analizar y detectar patrones en los datos recogidos en los entornos virtuales de aprendizaje<sup>1</sup> y aprovecharlos para mejorar el aprendizaje [24]. Cómo interactúan los estudiantes, cuándo lo hacen o con qué frecuencia consultan los recursos son cuestiones cuyas respuestas podrían utilizarse como indicadores de algunas competencias.

Decidir qué indicadores se pueden utilizar como evidencias de una u otra competencia es tarea de investigadores y docentes. A menos que se pueda considerar

---

<sup>1</sup>Las herramientas que dan soporte a los procesos de enseñanza-aprendizaje son conocidas como “entornos virtuales de aprendizaje”. Este término hace referencia a cualquier tipo de software educativo: LMS, wikis, mundos virtuales, etc. No ha de confundirse con el término VLE (Virtual Learning Environment) que se traduce también como “entorno de aprendizaje virtual”, pero que se utiliza también para referirse a los LMS [3]. Por tanto, en esta tesis se utilizará el término LMS para referirse a los cursos virtuales y se utilizará el término “entornos virtuales de aprendizaje” para referirse a cualquier tipo de software educativo. En ningún caso se utilizará el término VLE.



que la evaluación de una competencia genérica esté directamente alineada con la evaluación de una actividad específica en la herramienta, su evaluación requiere de la inventiva y originalidad del docente para ser capaz, por un lado, de diseñar actividades que obliguen al estudiante a desempeñar las competencias que se quieren evaluar y, por otro lado, diseñar evaluaciones que midan el desempeño del estudiante en dicha competencia a partir de la interacción de este con la herramienta. Ya que tanto los investigadores como los docentes se enfrentan a tareas de diseño, podemos decir que la educación se puede considerar como una *ciencia de diseño* [65].

### **La educación como una ciencia de diseño**

Cuando se trata de llevar a la práctica un método innovador diseñado por un investigador, este es incorporado en el sistema educativo y su contexto, de manera que su implementación final puede convertirse en algo muy diferente del diseño original. Los investigadores tienen que enfrentarse a la complejidad de las situaciones del mundo real y su resistencia al control experimental [28].

Para afrontar esta situación surgió una metodología de *investigación del diseño* (DBR, del inglés *Design-Based Research*) que no es clásicamente experimental, sino iterativa, refinando progresivamente el diseño inicial basado en la teoría conforme a la experiencia práctica. Según Hevner:

El DBR se basa en ideas procedentes de la base del conocimiento del dominio. La inspiración para la actividad de diseño creativo puede proceder de muy diversas fuentes para así incluir enriquecedores problemas u oportunidades desde entornos de aplicación, artefactos existentes, analogías/metáforas y teorías. Lo que se añade a la base del conocimiento como resultado de la investigación del diseño incluirá añadidos o extensiones de las teorías y métodos originales realizados durante la investigación, los nuevos artefactos (productos y procesos de diseño), y todas las experiencias ganadas desde el desempeño de los ciclos de diseño iterativos y pruebas sobre el campo de artefacto en el entorno de aplicación [56]

## 1. INTRODUCCIÓN

---

En esta tesis se pretende desarrollar un método para aplicar DBR en la evaluación de competencias genéricas. Para ello, se propondrá un método iterativo para el diseño de evaluaciones a partir de indicadores procedentes de los registros de las herramientas informáticas utilizadas por los estudiantes.

### 1.2 Contexto

La idea de que la educación pueda ser tratada como una *ciencia del diseño* viene de la década de los 90, con la ambición de llevar la investigación educativa de los laboratorios a la práctica. La enseñanza se considera una ciencia, pues los investigadores en educación investigan sobre ella, mientras que los docentes en general no investigan, sino que simplemente desarrollan y comparten teorías y explicaciones basadas en su propia experiencia [65]. Además, cuando un docente o investigador adopta un método innovador para implementarlo en sus clases, este es absorbido por el proceso normal de enseñanza, de forma que la implementación real puede convertirse en algo muy diferente del diseño original, es decir, hay un abismo entre la investigación y la práctica en la educación formal [4].

La metodología DBR fue concebida para solucionar esta separación entre la teoría y la práctica en la investigación. El DBR, cuyo método práctico principal es el experimento de diseño, es un enfoque de investigación mixto interdisciplinar que se lleva a cabo directamente en el área en la que se aplica y que enriquece también el conocimiento teórico de dicho área [87]. DBR no es una metodología clásicamente experimental, sino iterativa, que se basa en ir refinando progresivamente el diseño inicial basado en la teoría conforme a la experiencia práctica. Según el análisis realizado por Terry Andersen y Julie Shattuck [4], las características que un estudio DBR de calidad en la educación debe tener son las siguientes:

- *Estar situado en un contexto educativo real*: tener lugar en un contexto educativo real avala la validez de la investigación y asegura que los resultados puedan ser efectivamente utilizados para evaluar, informar y mejorar la práctica en, al menos, este contexto y probablemente en otros.
- *Enfocado en el diseño y prueba de una intervención significativa*: la selección y la creación de una intervención es una tarea colaborativa que atañe a

investigadores y docentes. La creación comienza con un preciso análisis del contexto local; se basa en la literatura relevante, en la teoría y en las prácticas de otros contextos; y se diseña específicamente para solventar un problema o aportar una mejora en la práctica. La intervención podría ser, por ejemplo, una actividad de aprendizaje, un tipo de evaluación, la introducción de una actividad administrativa (como un cambio en las vacaciones) o una intervención tecnológica.

- *Empleo de métodos mixtos*: las intervenciones DBR implican la aplicación conjunta de diferentes métodos mediante el empleo de una variedad de herramientas y técnicas de investigación. Los investigadores eligen, utilizan y combinan unos métodos u otros en función de sus necesidades.
- *Múltiples iteraciones*: la práctica del diseño suele implicar la creación y prueba de prototipos, refinamiento iterativo y la continua evolución del diseño, de la misma forma que ocurre en otros conocidos procesos de diseño como son, por ejemplo, la fabricación de coches o la moda.
- *Asociación colaborativa entre investigadores y docentes*: por una lado, los docentes suelen estar demasiado ocupados y no tienen experiencia para dirigir una investigación rigurosa. Por otro lado, los investigadores suelen carecer de conocimiento de la complejidad cultural, de la tecnología, de los objetivos y de las políticas de un sistema educativo que les permita crear y medir eficientemente el impacto de una intervención. Por tanto, se requiere una asociación para el estudio.
- *Evolución de los principios de diseño*: el diseño evoluciona desde y hacia la elaboración de principios de diseño, patrones y teorías funcionales. Estos principios no son diseñados para crear fundamentos o teorías que tengan el mismo efecto en cualquier contexto, sino que sirven para ayudarnos en la comprensión del contexto y la intervención, y nos ayuden para ajustar ambos y así maximizar el aprendizaje. El desarrollo de principios prácticos de diseño es una parte fundamental del DBR, y pone en desventaja a aquellos

## 1. INTRODUCCIÓN

---

tipos de investigación que unilateralmente comienzan con las pruebas en clase y después desaparecen con el investigador una vez que el experimento ha concluido.

- *Comparación con la investigación-acción:* tanto los docentes como los investigadores encuentran a menudo confuso diferenciar entre DBR e investigación-acción. Sin embargo, aunque ambas metodologías se sitúan dentro del campo de la investigación aplicada, difieren en características principales. Mientras que la investigación-acción se concibe principalmente para alcanzar una serie de objetivos a nivel local, en DBR se pretende también evolucionar a nivel teórico, maximizando la generalización y el entendimiento en la comprensión de aplicaciones prácticas. Además, la investigación-acción es llevada a cabo normalmente por un solo docente, por lo que no se beneficia de la experiencia y la energía que caracterizan a los equipos de investigación y diseño DBR.
- *Repercusión en las prácticas:* el DBR no debe avanzar únicamente en el campo teórico, sino que para demostrar y justificar su valor real deberá ser además implementado en un contexto de estudio local.

### 1.3 Objetivos y preguntas de investigación

El principal objetivo de esta tesis es:

**Proporcionar un método y una herramienta informática para diseñar y contrastar estrategias de evaluación de competencias genéricas a partir de los registros de actividad de los entornos virtuales de aprendizaje.**

#### **Preguntas de investigación**

Para alcanzar dicho objetivo, se comenzará definiendo una serie de preguntas de investigación a las que se tratará de dar respuesta mediante una revisión de la literatura. Las preguntas de investigación son las siguientes:

### **1.3 Objetivos y preguntas de investigación**

---

**Q1. ¿Qué competencias se han evaluado de forma automática o asistida por ordenador a partir de la actividad de los estudiantes en los entornos virtuales de aprendizaje?**

La evaluación de competencias genéricas no es una cuestión reciente, y son muchas las que han sido evaluadas a lo largo de los años. De hecho, se pueden encontrar cientos de trabajos que abordan su evaluación en la literatura. Sin embargo, para esta revisión nos centraremos en aquellos que, partiendo de la actividad de los estudiantes en entornos virtuales de aprendizaje, busquen facilitar la labor del docente mediante el uso de herramientas informáticas que asisten o automatizan el proceso.

**Q2. ¿Qué métodos se utilizan para evaluar competencias genéricas en entornos virtuales de aprendizaje?**

Una de las contribuciones de esta tesis es un método para la evaluación de competencias genéricas en entornos virtuales de aprendizaje. Pero antes debemos conocer y valorar qué métodos se han estado utilizando hasta ahora. Para ello se recopilarán los métodos de evaluación que se han empleado para evaluar cada una de las competencias genéricas obtenidas en la pregunta anterior.

**Q3. ¿Qué técnicas se utilizan para evaluar competencias genéricas a partir de los registros de actividad de un entorno virtual de aprendizaje?**

Por último, deberemos responder a la pregunta de qué técnicas se han empleado para implementar los métodos que se han utilizado para evaluar las competencias genéricas en entornos virtuales de aprendizaje.

#### **Objetivos específicos**

A partir del objetivo principal de esta tesis de proporcionar un método a los docentes para diseñar y contrastar estrategias de evaluación de competencias genéricas a partir de los registros de actividad de los entornos virtuales de aprendizaje, así como una herramienta informática que implemente el método, se obtienen dos objetivos específicos:

## 1. INTRODUCCIÓN

---

### **O1. Definir un método que permita al docente obtener de manera automática un conjunto de indicadores de los entornos virtuales de aprendizaje**

Basándonos en la metodología DBR se define el método de evaluación denominado *design-based assessment* (DBA), un método que permitirá evaluar las competencias genéricas de los estudiantes a partir de su actividad en los entornos virtuales de aprendizaje.

### **O2. Definir un lenguaje específico de dominio (DSL) que permita a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje**

Se definirá un lenguaje específico de dominio (DSL, del inglés *Domain-Specific Language*) para poner en práctica el método del objetivo O1 que sea aplicable a diferentes entornos virtuales de aprendizaje. También se implementarán las herramientas informáticas necesarias para interpretar las consultas del DSL.

### **Hipótesis de investigación**

Para evaluar la consecución de los objetivos propuestos se plantean las siguientes hipótesis:

#### **H1: El método DBA permite obtener de manera automática indicadores de los entornos virtuales de aprendizaje.**

Esta hipótesis se corresponde con el objetivo *O1* y persigue conocer si el método propuesto permite obtener de manera automática los valores de un conjunto de indicadores previamente definidos por el diseñador de la evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje.

#### **H2: El DSL permite a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje.**

Esta hipótesis se corresponde con el objetivo *O2* y persigue evaluar si el DSL permite a los docentes diseñar y contrastar estrategias de evaluación a partir de los

registros de actividad de los entornos virtuales de aprendizaje.

### 1.4 Estrategia de investigación

En este apartado se describe y justifica el uso de la estrategia de investigación llevada a cabo en esta tesis doctoral.

#### Diseño y creación

La estrategia de investigación a utilizar debía contemplar como contribución a la ciencia el desarrollo de herramientas informáticas. Ante este requisito, se decidió utilizar la estrategia de investigación *Design and Creation* (Diseño y Creación) de Oates, que se basa en el desarrollo de un nuevo artefacto o producto tecnológico [75]. Los tipos de artefactos que abarca esta estrategia son constructores (*constructs*), modelos (*models*), métodos (*methods*) e instancias (*instantiations*). En alguna ocasión, como ocurre en esta tesis, puede haber más de una contribución.

Las dos contribuciones de esta tesis son:

1. Un método para aplicar la ciencia del diseño en la evaluación de competencias genéricas de los estudiantes a partir de indicadores procedentes de los registros de actividades de aprendizaje
2. Herramientas informáticas que automaticen y den soporte a la aplicación del método.

Al ser el propio método la principal contribución de esta investigación, la estrategia de *Diseño y creación* no necesita ser combinada con ninguna otra estrategia. La estrategia de diseño y creación se basa en los principios establecidos del desarrollo de sistemas, siendo un enfoque típico de resolución de problemas que utiliza un proceso iterativo de 5 pasos que serán abordados en los diferentes capítulos de esta tesis: conocimiento (*awareness*), recomendación (*suggestion*), desarrollo (*development*), evaluación (*evaluation*) y conclusión (*conclusion*).

- Conocimiento, presentado como el *estado del arte* (capítulo 2), consiste en el reconocimiento y articulación del problema a ser estudiado a partir de la literatura. Para abordarlo se realizó un *estudio sistemático de mapeo* que pudiera

## 1. INTRODUCCIÓN

---

dar respuesta a las preguntas de investigación presentadas en el apartado anterior.

- Recomendación, presentado como el *resumen de problemas encontrados* (capítulo 3), es donde se recapitulan los principales métodos y técnicas informáticas utilizados en los trabajos recogidos en el estado del arte y se ofrece una nueva idea de cómo el problema podría ser abordado.
- Desarrollo, presentado como el *método para la evaluación de competencias genéricas* (capítulo 4), describe tanto el método como las herramientas informáticas que se presentan como principales contribuciones de esta tesis.
- Evaluación, presentado también como *evaluación* (capítulo 5), examina el desarrollo del método y se busca una evaluación de su valor y su desviación de las expectativas.
- Conclusión, presentado como *conclusiones* (capítulo 6), es donde los resultados desde el proceso de diseño son consolidados y criticados, se indica el conocimiento obtenido, junto con los resultados inesperados o anómalos que no pudiesen ser aún explicados y que podrían ser el objeto de futuras investigaciones.



## Estado del Arte

Las competencias genéricas son las habilidades que los profesionales deben ser capaces de desempeñar independientemente de su especialización. Habilidades como el trabajo en equipo, la comunicación interpersonal, la capacidad para resolver problemas, la creatividad o el liderazgo, entre otras, son competencias que las empresas demandan hoy en día en los nuevos titulados como complemento de las competencias específicas que se les supone por la titulación que hayan estudiado. Desde un punto de vista formativo, los docentes deben integrar estas competencias en los procesos educativos, tanto en las clases tradicionales como en los entornos virtuales de aprendizaje. Y por supuesto, deben fijar mecanismos no sólo para el desarrollo de estas competencias, sino también para la evaluación de las mismas.

Un LMS es un tipo de entorno virtual de aprendizaje que almacena información de estudiantes, docentes, cursos, tareas, trabajos, etc. Estos elementos se relacionan y configuran para ofrecer al usuario una experiencia de curso virtual. Estos cursos virtuales están muy extendidos hoy en día, siendo el soporte virtual de las clases presenciales o incluso siendo el único medio donde unas clases o un curso se imparten. Las clases virtuales presentan numerosas ventajas con respecto a las clases tradicionales. Por un lado se elimina la limitación geográfica que tienen las clases tradicionales, y por otro lado la oferta y variedad de cursos ofrecidos suele ser más amplia. Además, para los estudiantes presentan otras ventajas fundamentales: en

## 2. ESTADO DEL ARTE

---

primer lugar la flexibilidad de horario, permitiéndoles compatibilizar los estudios con una vida laboral sin renunciar a crecer profesionalmente; y en segundo lugar, les permite estar en contacto permanente con otros estudiantes y docentes mediante diferentes herramientas (foros, chats, etc.) [2].

Pero además de todo lo anterior, un entorno virtual de aprendizaje almacena una gran cantidad de información que adecuadamente presentada puede ser de gran utilidad para los docentes para analizar el trabajo de sus estudiantes [82]. Cada archivo, cada acceso o cada tarea realizada por los estudiantes queda registrada en el sistema. Por desgracia, esta información no está siempre a disposición del docente, y si lo está, requiere un filtrado para poder ser utilizada [25]. Según el informe *Making Data Work For Teachers and Students*, de la *Fundación Bill & Melinda Gates* [51], el 67% de los docentes no están del todo satisfechos con el aprovechamiento que sacan de los datos y las herramientas que utilizan regularmente. Además, en este mismo informe se extrae que el 78% de los docentes cree que los datos les ayudarían a validar lo que sus estudiantes son y lo que pueden conseguir.

La información almacenada en el registro de un entorno virtual de aprendizaje podría darnos indicadores sobre cómo actuarían los estudiantes en situaciones laborales reales y podría utilizarse para medir cómo de eficaz es un estudiante trabajando en equipo, cómo se planifica un estudiante a la hora de realizar sus tareas o si lidera eficazmente un grupo. ¿Podríamos entonces afirmar que los registros de actividad de un entorno virtual de aprendizaje pueden utilizarse para evaluar competencias genéricas? Para responder a esta cuestión y establecer la base teórica sobre la que se sustenta esta tesis doctoral se presenta un *estudio de mapeo sistemático* (SMS, del inglés *Systematic Mapping Study*). Un SMS es una amplia revisión de los estudios primarios en un área específica cuyo objetivo es identificar alguna evidencia sobre el tema y cuyo punto de partida inicial serán las preguntas de investigación que se definirán en el siguiente apartado.

### 2.1 Preguntas de investigación del SMS

El objetivo principal de esta tesis doctoral (presentado en la sección 1.3) es:

*Proponer un método para evaluar a los estudiantes en el desempeño de sus competencias genéricas mediante indicadores procedentes de los registros de actividades de aprendizaje.*

Para abordar este objetivo ha de conocerse primero el estado del arte, dando respuesta para ello a diferentes preguntas de investigación tales como cuáles son las competencias genéricas que se han evaluado haciendo uso de métodos y técnicas informáticas, cuáles son esos métodos y si se están usando para este fin los registros de actividad de los entornos virtuales de aprendizaje.

Por tanto, partiendo del objetivo principal, se definen las siguientes preguntas de investigación para el SMS:

- SMS1. ¿Qué competencias se han evaluado de forma automática o asistida por ordenador a partir de la actividad de los estudiantes en los entornos virtuales de aprendizaje?
- SMS2. ¿Qué métodos se utilizan para evaluar competencias genéricas en entornos virtuales de aprendizaje?
- SMS3. ¿Qué técnicas se utilizan para evaluar competencias genéricas a partir de los registros de actividad de un entorno virtual de aprendizaje?

## 2.2 Metodología

Un SMS es una amplia revisión de los estudios primarios en un área específica cuyo objetivo es identificar alguna evidencia sobre el tema. Este estudio se basa en las directrices de la metodología propuesta por Kitchenham [62] sobre cómo se deben planificar, ejecutar y presentar los resultados de una revisión de la literatura en ingeniería del software. Y en particular, para este trabajo se ha utilizado la propuesta de Petersen [80], que describe una serie de etapas con las que llevar a cabo una revisión de la literatura a partir de un mapeo sistemático. Se comenzará describiendo un protocolo de revisión, los motores de búsqueda y los términos de búsqueda a emplear.

## 2. ESTADO DEL ARTE

---

### Protocolo de revisión

La definición del protocolo de revisión requiere la especificación de una serie de pasos para obtener la bibliografía de nuestro estudio. Los pasos a seguir son los siguientes:

1. Selección de motores de búsqueda (sección 2.2).
2. Definición de los términos de búsqueda (sección 2.2).
3. Determinación de los criterios de selección (sección 2.2.1).
4. Clasificación para la extracción de datos (sección 2.2.2).

### Motores de búsqueda

Para encontrar la bibliografía, se realizarán consultas en las siguientes bibliotecas digitales:

- Web of Science
- Wiley Online Library
- Science Direct
- IEEE Digital Library (Xplore)

### Términos de búsqueda

Existen diversos términos que pueden utilizarse para referirse a la evaluación de competencias genéricas de manera automatizada o asistida. Por la naturaleza de nuestro trabajo, debemos contemplar siempre en las palabras de búsqueda los términos *assessment* y *generic skills* o *generic competences*. Realizar la búsqueda por el término *Assessment of generic skills* o *assessing generic skills* devolvía muy pocos resultados. Por ejemplo, en la *Wiley Online Library* la búsqueda del término exacto *generic skills assessment* devolvió un único resultado. Sin embargo, debilitar la búsqueda con términos como *generic competences* o *generic skills* junto con la palabra *assessment* daba un número de resultados muy elevado. En la misma biblioteca, buscar por los términos “*generic skills*” *and student and assessment* nos

FUENTE	TÉRMINOS DE BÚSQUEDA	PUBLICACIÓN	RESULTADOS
Web of Science	((“generic competences“ OR “generic skills“) AND assessment)	Journals	138
Wiley Online Library	“generic competences“ AND assessment	Journals and Conferences	50
Science Direct	(“generic competences“) AND assessment)	Journals	71
IEEE Digital Library (Xplore)	((“generic competences“) AND assessment)	Journals and Conferences	54
TOTAL			313

**Tabla 2.1:** Resumen de búsqueda de bibliografía

devolvía 609 resultados. En primera instancia se probó añadiendo términos como *E-Learning*, *computer-assisted* o *mobile learning*. Sin embargo, incluir términos de este tipo reducía también drásticamente el número de resultados obtenidos en la búsqueda, no llegando a obtenerse bibliografía más significativa que si no se incluían. Por tanto, a tenor de las pruebas se decide eliminar de la búsqueda ese tipo de términos. La combinación de los términos de búsqueda empleados en la investigación, así como a los motores de búsqueda que fueron aplicados en cada una pueden observarse en la tabla 2.1. Los términos de búsqueda se han empleado en todos los campos (título, resumen, texto, etc.).

### 2.2.1 Criterios de selección

Para determinar si un trabajo debía formar parte de nuestra selección de estudios primarios se leyó el título, el resumen y las palabras clave. Cuando esto no era suficiente se complementaba la lectura anterior con una somera la lectura del artículo completo, y más detallada de la introducción y las conclusiones. Nuestra búsqueda se centró en la localización de los trabajos que, habiendo sido obtenidos en el proceso de búsqueda anterior, vayan en línea con nuestro estudio y puedan ayudarnos

## 2. ESTADO DEL ARTE

---

a resolver las preguntas de investigación. Para ello, se realizó la proyección de los trabajos seleccionados utilizando los siguientes criterios de exclusión:

- Includido (*included*): trabajo relacionado con nuestra investigación.
- Fuera del tema (*off Topic*): trabajo no relacionado directamente con nuestra investigación. Son trabajos que satisfacen los criterios de búsqueda, pero cuya contribución no está directamente relacionada con la temática de este estudio. La mayoría de artículos descartados en este bloque consisten en experiencias que trabajan o mejoran alguna competencia genérica en los estudiantes, pero no mencionan si después el desempeño en la competencia se mide de alguna forma, y si por el contrario sí realizan una medición, lo hacen sin apoyo alguno de la tecnología.
- Idioma no disponible (*Unsupported Language*): trabajo escrito en un idioma diferente al inglés o español. La mayoría de los textos son en inglés, por lo que este criterio de descarte apenas es utilizado.
- Duplicado (*duplicated*): trabajos cuya contribución principal está recogida en otros trabajos ya incluidos.
- No leído (*unread*): trabajo que no ha podido ser leído. Son textos que no han sido leídos al no estar disponible en las bibliotecas digitales a las que se tiene acceso desde la Universidad de Cádiz ni se ha podido encontrar por otros medios (petición por correo a los autores, búsqueda en repositorios abiertos de Internet, etc.).

### 2.2.2 Esquema para la extracción de datos

Para la extracción de la información se han dividido los trabajos de acuerdo a los siguientes tres aspectos: tipo de investigación, tipo de contribución y ámbito de aplicación de la investigación. A continuación se detalla esta clasificación.

### Tipo de investigación

Esta clasificación hace referencia al tipo de trabajo de investigación llevado a cabo por los investigadores. Existen diferentes enfoques para la clasificación de los trabajos según el tipo investigación que desarrollan. Algunos de estos sistemas de clasificación son los propuestos por Wieringa [103] y Hevner [57]. Usamos el primero, ya que es el recomendado en el SMS descrito por Petersen [80].

- Solución propuesta (*proposal of solution*): se propone una solución para un problema; la solución puede ser innovadora o una extensión significativa de una técnica existente. Los posibles beneficios y la aplicabilidad de la solución se demuestran por un pequeño ejemplo o una buena línea de argumentación.
- Investigación por validación (*validation research*): las técnicas investigadas son nuevas y todavía no se han aplicado en la práctica. Estas técnicas podrían ser por ejemplo los experimentos, es decir, el trabajo realizado en un laboratorio.
- Investigación por evaluación (*evaluation research*): las técnicas se aplican en la práctica y se lleva a cabo una evaluación de la técnica. Se muestra cómo se implementa la técnica en la práctica (implementación de la solución) y cuáles son las consecuencias de la aplicación en términos de ventajas y desventajas (evaluación de implementación).
- Artículos de experiencia (*experience papers*): trabajos que explican qué y cómo algo se ha llevado a cabo en la práctica. Basado en la experiencia personal de los autores.
- Artículos de opinión (*opinion papers*): estos trabajos expresan la opinión personal de alguien acerca de la bondad o viabilidad de una determinada técnica, o sobre cómo se debería implementar dicha técnica.
- Trabajos filosóficos (*philosophical papers*): estos trabajos esbozan una nueva forma de ver las teorías existentes, estructurando el campo en forma de una taxonomía o un marco conceptual.

## 2. ESTADO DEL ARTE

---

### Tipo de contribución

En este apartado se clasifican los trabajos según el tipo de contribución que realizan estos al ámbito en el que se desarrollan. Una vez realizado el estudio sistemático de la literatura y habiendo seleccionado los artículos, se realiza una clasificación en base a su aportación.

Lamentablemente, el uso de algunos términos puede ser confuso, debido a las diversas interpretaciones del mismo que hacen distintos autores. Algunos de estos términos son framework, modelo, estrategia, proceso, procedimiento, método o metodología. Nuestra interpretación es la siguiente:

- Modelo (*model*): es una representación de procesos, modelos o sistemas pertenecientes a un supra-sistema, cuyo fin es el análisis de interacción de ellos para mantener una relación flexible que les permita cumplir su función particular y cumplir la función de dicho supra-sistema.
- Método (*method*): contempla aquellos trabajos cuya contribución sea descrita por los autores como una serie de pasos.
- Herramienta (*tool*): se utiliza para los artículos que presentan un software independiente o una extensión de algún otro programa.
- Framework (*framework*): aquí se consideran aquellos trabajos que contribuyen con una combinación de los elementos anteriores (es decir, con un modelo, un proceso y una herramienta).
- Técnica (*technique*): un procedimiento utilizado para llevar a cabo una actividad o tarea específica. Podría venir acompañado de una herramienta de apoyo.

### Ámbito de aplicación de la investigación

Además de las clasificaciones anteriores, es necesario extraer más información acerca de los conceptos que representan la contribución de investigación. Para ello se recopila información sobre el ámbito y la manera en la que se realiza la evaluación de competencias en cada contribución. Una vez identificada esta información, se agrupan según sus similitudes, quedando finalmente la siguiente clasificación:



- Evaluación del docente (*teacher assessment*): el docente evalúa el desempeño de los estudiantes en una o varias competencias genéricas de manera asistida o semi-asistida por un sistema informático.
- Evaluación entre iguales y autoevaluación (*peer and self-assessment*): en estos trabajos los estudiantes se encargan de todo o de parte del proceso de evaluación, ya sea evaluando su propio trabajo (autoevaluación) o el de sus compañeros (evaluación entre iguales), y utilizando para ello algún tipo de sistema informático. Los investigadores animan a los docentes a utilizar este tipo de evaluación ya que ayuda a los estudiantes a reflexionar sobre su propio aprendizaje y favorecer el aprendizaje colaborativo en el aula [26].
- Herramientas de evaluación semiautomática (*semiautomatic assessment tools*): en esta categoría se recogen trabajos que automatizan en parte el proceso de evaluación de competencias.

### 2.2.3 Visualización y análisis de los datos

Tras obtener los estudios primarios, hay una etapa de análisis, donde se resumen los datos extraídos para responder a las preguntas de investigación planteadas. El análisis de los resultados se centra en el estudio de las publicaciones para cada categoría y por lo tanto, en la determinación del grado de cobertura de cada categoría. Esta información generalmente se resume en tablas y gráficos. Otro método utilizado en nuestro estudio es la combinación de diferentes categorías (por ejemplo, el ámbito de investigación contra el tipo contribución) y su representación se muestra en un mapa sistemático en forma de gráfico de burbujas.

En el siguiente capítulo se presentan los resultados obtenidos.

## 2.3 Resultados

A continuación se muestran los resultados del estudio. Comienza el capítulo con la localización de los estudios primarios, para continuar con la extracción de los

## 2. ESTADO DEL ARTE

---

datos de estudio, mostrándose varios gráficos y tablas que justifican la información mostrada. Finalmente se categorizan los estudios y se muestra el esquema de clasificación resultante.

### 2.3.1 Localización de la literatura

En la tabla 2.1 se muestran las búsquedas realizadas en las bibliotecas digitales más importantes en ingeniería informática, los términos de búsqueda utilizados y el número de documentos obtenidos. En cada biblioteca, se utilizaron los formularios de búsqueda avanzada y los resultados fueron obtenidos a fecha 21 de agosto de 2015. Toda la información resultante de las búsquedas de este SMS está disponible para su consulta <sup>1</sup>.

En total se recopilaron 313 trabajos. El número de estudios primarios resultante (después de aplicar criterios de selección y exclusión) fue de 30 trabajos (menos de un 10% del total de trabajos recopilados). Además, 268 trabajos se catalogaron como “fuera del tema”. Los resultados de esta clasificación pueden verse en la tabla 2.2. A tenor de los resultados ofrecidos, y sobre todo del número de trabajos descartados, el lector puede pensar que se debieron añadir términos informáticos entre los criterios de búsqueda. Sin embargo, el número de términos informáticos que se podrían añadir a las búsquedas era muy amplio y determinados trabajos que han sido recopilados en este trabajo hubieran quedado fuera en ese caso porque hubiera sido imposible tener todos los términos en cuenta. Además, el hecho de que las bases de datos utilizadas sean específicas de informática y computación es garantía suficiente de que sus aportaciones deben ser en su mayoría basadas en tecnología.

### 2.3.2 Extracción de datos

Aunque las tecnologías entraron a formar parte de la vida académica hace ya varios años, no es hasta 2013, con la tercera generación de herramientas de medición educativa según el marco de la Comisión Europea (*Generation 3: continuous integrated assessment*) [86], cuando se comienzan a integrar la evaluación en las

---

<sup>1</sup><https://goo.gl/rzTCHN>

## 2.3 Resultados

CRITERIO	TRABAJOS	PORCENTAJE
Incluido	30	9,58%
Fuera del tema	268	85,62%
Idioma no disponible	0	0,00%
Duplicado	10	3,20%
No leído	5	1,60%
TOTAL	313	100,00%

**Tabla 2.2:** Clasificación de trabajos tras aplicar los criterios de selección y exclusión

AÑOS	RESULTADOS	PORCENTAJE
2007	2	7%
2008	2	7%
2009	2	7%
2010	3	10%
2011	4	13%
2012	1	3%
2013	9	30%
2014	4	13%
2015	3	10%

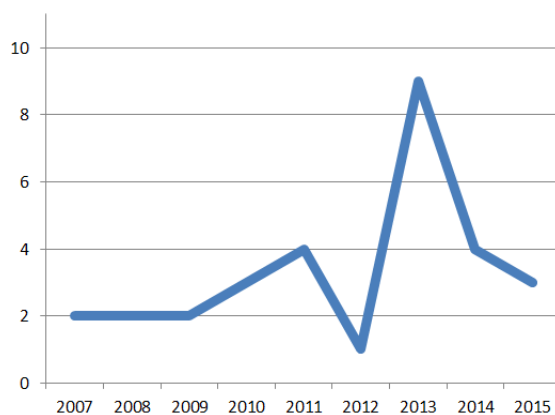
**Tabla 2.3:** Cantidad de trabajos publicados cada año

herramientas de aprendizaje. Es a partir de entonces cuando conceptos como *Data Mining and analysis*, *Behavioural tracking* and *Learning analytics* comienzan a usarse. Tanto en la tabla 2.3 como en la figura 2.1 puede verse la distribución de la producción de la selección primaria a lo largo de los años. La mayoría de los seleccionados se pueden localizar en los últimos años. Véase como 16 de estos trabajos (53,33%) fueron publicados entre 2013 y 2015.

Todos los trabajos seleccionados evalúan una o varias competencias genéricas utilizando algún método, técnica e instrumento de evaluación. Para el esquema de clasificación se concreta para cada trabajo las competencias que evalúa y con

## 2. ESTADO DEL ARTE

---



**Figura 2.1:** Distribución de las publicaciones por años

qué método, técnica y herramienta lo hace. A continuación se describen para cada una de estas características los valores que aparecerán en la revisión para que queden definidos todos los términos que aparecen en el capítulo del esquema de clasificación.

### **Competencias genéricas**

En la tabla 2.4 se muestran las competencias genéricas que se evalúan en los trabajos seleccionados. De esta manera, se unifica la denominación de las competencias, ya que en ocasiones los autores se refieren a las mismas competencias de diferentes denominaciones. En la primera columna se muestra la denominación corta que se utilizará para referirnos a cada competencia. Al clasificar los artículos se han utilizado las competencias genéricas definidas en el *Tuning Educational Structures in Europe* [52].

COMPETENCIA	DESCRIPCIÓN
Análisis	Capacidad de abstracción, análisis y síntesis [52].
Aprendizaje permanente	Marco constituido por el aprendizaje formal, no formal e informal, que aspira a la adquisición de conocimiento para alcanzar el máximo desarrollo de la personalidad y de las destrezas profesionales en las diferentes etapas de la vida [99].
Comunicación	Habilidad para comunicarse de manera tanto oral como escrita en la lengua materna [52].
Creatividad	Capacidad para crear nuevas ideas [52].
Cultural	Aprecio y respeto por la diversidad y la multiculturalidad [52].
Emprendimiento	Capacidad para tomar la iniciativa y espíritu de empresa [52].
Gestión de proyectos	Habilidades para diseñar y gestionar proyectos [52].
Habilidades interpersonales	Capacidad de la persona para comunicarse e interactuar con otras personas [52].
Investigación	Capacidad para llevar a cabo la investigación en un nivel apropiado [52].
Lengua extranjera	Capacidad de los estudiantes para comunicar sus ideas en un segundo idioma [50].
Liderazgo	Habilidad para motivar a la gente y conducirlos hacia un objetivo común [52].
Pensamiento crítico	Habilidad para interpretar, analizar y evaluar ideas y argumentos [46].
Planificación y gestión del tiempo	Capacidad de planificar y gestionar el tiempo de manera efectiva [52].
Resolución de problemas	Habilidad para identificar, plantear y resolver problemas [52].
Responsabilidad	Capacidad para actuar con responsabilidad social y conciencia cívica [52].
TIC	Habilidades en el uso de las tecnologías de la información y de la comunicación [52].
Toma de decisiones	Capacidad para tomar decisiones razonadas [52].
Trabajo autónomo	Capacidad para trabajar de forma autónoma [52].
Trabajo en equipo	Trabajo realizado por un conjunto de personas en aras de un objetivo común [35].

**Tabla 2.4:** Competencias genéricas

## 2. ESTADO DEL ARTE

---

### Métodos

En cada uno de los trabajos seleccionados se llevaba a cabo un método de evaluación. A partir de la descripción que realizaban los autores sobre el proceso de evaluación realizado, cada trabajo ha sido clasificado dentro de alguno de los métodos que a continuación se detallan:

- Evaluación formativa (*formative assessment*): proceso utilizado por docentes y estudiantes para identificar y reaccionar al aprendizaje de los estudiantes con el fin de mejorar dicho aprendizaje mientras éste tiene lugar [16]. Es una parte integral del proceso educativo cuyo objetivo es proporcionar información de manera sistemática y continua sobre el propio proceso.
- Evaluación sumativa (*summative assessment*): método que mide el grado de éxito de los estudiantes a la hora de alcanzar los criterios utilizados para medir los objetivos de aprendizaje fijados para el curso o módulo. Se utiliza para cuantificar los logros y proporcionar información para la aptitud o no del evaluado de pasar al siguiente nivel o módulo [100].
- Evaluación auténtica (*authentic assessment*): evaluación en la que las tareas están estrechamente alineadas con lo que el estudiante experimentará en el mundo laboral. Este tipo de evaluación está diseñado para desarrollar en los estudiantes habilidades y competencias en paralelo al desarrollo académico [55].
- Evaluación diagnóstica (*diagnostic assessment*): al igual que la evaluación formativa, es una evaluación que pretende mejorar la experiencia del estudiante y su nivel de éxito. La diferencia con respecto a la evaluación formativa radica en que mientras que la evaluación diagnóstica se realiza antes de que comience el proceso educativo con la intención de conocer lo que el estudiante ya sabe y las dificultades que se encontrará, la formativa se realiza durante el proceso [60].
- Evaluación dinámica (*dynamic assessment*): mide lo que el estudiante logra cuando se le imparten algunas nociones sobre un tema o campo desconocido. Puede ser útil para evaluar el potencial en un tema concreto en ausencia de

un logro anterior relevante, o para evaluar el potencial de aprendizaje general para los estudiantes que tienen un contexto desfavorecido. A menudo se utiliza antes de que el cuerpo principal de la enseñanza [66].

- Evaluación sinóptica (*synoptic assessment*): estimula a los estudiantes a combinar elementos de su aprendizaje desde diferentes partes de un programa para mostrar el conocimiento acumulado y la comprensión de un tema o materia específica. Una evaluación sinóptica normalmente permite a los estudiantes mostrar su capacidad de integrar y aplicar sus habilidades, conocimientos y entendimiento con amplitud y profundidad en el tema. Puede ayudar a evaluar la capacidad de los estudiantes de aplicar el conocimiento y la comprensión obtenida en una parte del programa para mejorar sus comprensión en otras partes del programa o a lo largo del mismo. La evaluación sinóptica puede formar parte de otras formas de evaluación [83].
- Evaluación con referencia al criterio (*criterion referenced assessment*): los logros de cada estudiante se juzgan con respecto a criterios específicos, sin tener en cuenta lo logrado por otros estudiantes. En la práctica, la comparación con otros sujetos puede afectar al enjuiciar si un criterio específico se ha cumplido o no [38].
- Evaluación ipsativa (*ipsative assessment*): esta evaluación se realiza con respecto al nivel previo del estudiante. Puede medir el nivel con el que un estudiante ha desempeñado una tarea particular con respecto a su propio nivel medio de desempeño, con respecto a su mejor trabajo o con respecto a su trabajo más reciente. La evaluación ipsativas se suele correlacionar con el esfuerzo, para promover la recompensa al trabajo y mejorar la motivación para aprender [59].

### Técnicas

En cada uno de los trabajos seleccionados se utilizó alguna técnica para llevar a cabo alguno de los métodos de evaluación anteriormente descritos. A continuación se describen el conjunto de técnicas que aparecen en estos trabajos:

## 2. ESTADO DEL ARTE

---

- *Observación sistemática*: técnica para la recolección de datos sobre el aprendizaje basada en inspección y estudio esencialmente descriptivo de la actividad, evento o hecho realizado por el estudiante a evaluar. Algunos instrumentos típicos son rúbricas, guías de observación y listas de comprobación.
- *Pruebas escritas*: instrumento de medición cuyo propósito es que el estudiante demuestre la adquisición de un aprendizaje cognitivo, o el desarrollo progresivo de una destreza o habilidad. Por sus características, requiere contestación escrita por parte del estudiante [90]. Algunos instrumentos que se utilizan en esta técnica son los exámenes escritos y los cuestionarios.
- *Pruebas orales*: interacción oral evaluador-evaluado mediante la que el evaluado busca acreditar conocimiento sobre un tema determinado ante uno o varios evaluadores que a vez utilizarán esta exposición para calificar al evaluado en el desempeño de alguna competencia.
- *Indicadores basados en logros*: indicadores que marcan si un resultado final o logro se ha alcanzado y para determinar así el grado en que cada competencias ha sido desarrollada. El propósito es conseguir que los estudiantes desarrollen competencias, y los indicadores son el recurso para evaluar dicho desarrollo. Esta técnica es típica en los juegos serios, en donde las distintas fases o etapas del juego con sus correspondientes logros se mapean a indicadores del nivel de desempeño de competencias [43].
- *Indicadores de trabajo en entornos virtuales de aprendizaje*: indicadores del proceso de desarrollo del trabajo realizado por los estudiantes (no del trabajo final en sí). Encontramos estas técnicas en experimentos en los que tras haber trabajado los estudiantes en diferentes actividades del entorno virtual de aprendizaje son evaluados del desempeño en diferentes competencias a partir de los registros de interacción de estos estudiantes con el propio entorno.
- *Actividades sin determinar*: se utiliza este término para referirnos a las técnicas que se han utilizado en trabajos en los que se indica que se han evaluado actividades pero no se menciona explícitamente el cómo. Por contexto pueden ser tanto actividades del entorno virtual de aprendizaje como actividades



presenciales. No se presta atención a cómo son y cómo se evalúan esas actividades en particular, sino a que después se toma la calificación de cada actividad y de una u otra forma se mapea a la evaluación de alguna competencia. Quizás se usen rúbricas, cuestionarios o entrevistas, pero o no se mencionan o si lo hacen no está directamente relacionado con la evaluación de la competencia genérica.

### **Instrumentos de evaluación**

Los instrumentos de evaluación son las herramientas que se utilizan para llevar a cabo las técnicas de evaluación. En el siguiente listado se muestran los instrumentos de evaluación utilizados en los trabajos seleccionados en el apartado anterior junto con su descripción:

- La *rúbrica* es un instrumento de evaluación basado en una escala cuantitativa y/o cualitativa asociada a unos criterios preestablecidos que miden las acciones del alumnado sobre los aspectos de la tarea o actividad que serán evaluados. Básicamente, existen dos grupos: las holísticas, que tratan de evaluar el aprendizaje o competencia desde una visión más global, y las analíticas, que se centran en algún área concreta de aprendizaje [97].
- La *Entrevista* es una prueba oral en la que el estudiante debe desarrollar el tema que el docente le indique y/o responder a las preguntas que éste le formule. Este método se ha utilizado en algunos trabajos como [102] para que los alumnos justifiquen de forma razonada las respuestas que dieron a las preguntas de evaluación.
- El *cuestionario* consiste en un conjunto de preguntas preparado sistemática y cuidadosamente, sobre los hechos y aspectos que interesan en la evaluación. Es una técnica de evaluación que puede abarcar aspectos cuantitativos y cualitativos. Su característica singular radica en que para registrar la información solicitada a los mismos sujetos, se usa una forma menos profunda e impersonal, que el “cara a cara” de la entrevista [74].

## 2. ESTADO DEL ARTE

---

- La *herramienta de seguimiento* es una herramienta para monitorizar el trabajo del estudiante a lo largo del semestre. En el trabajo presentado en [63] el docente utilizó un diario para anotar la evolución de cada estudiante.
- *Tests automáticos*: algunos de los tests automáticos que se han encontrado dentro de la bibliografía son tests de personalidad. Este tipo de test está diseñado para revelar aspectos del carácter o mecanismos psicológicos de un individuo. La evaluación de la personalidad se puede ver como la aplicación de procedimientos para medir aspectos de la personalidad de manera que sean aplicables a otros dominios [104]. Uno de esos dominios es el laboral, sobre todo las entrevistas de trabajo. Es común la necesidad del empresario de conocer la aptitud del candidato a un puesto para asumir cierto rol dentro de una empresa. Todas las competencias están relacionadas por tanto con características de los encuestados que sean de interés para los empleadores (trabajo en equipo, responsabilidad, comunicación, habilidades interpersonales, creatividad, gestión de proyectos, liderazgo, resolución de problemas, etc.).
- *Juegos serios (Serious games)*: son juegos diseñados para un propósito principal distinto del de la pura diversión [36]. Normalmente, el adjetivo “serio” pretende referirse a productos utilizados por industrias como la de defensa, educación, exploración científica, sanitaria, urgencias, planificación cívica, ingeniería, religión y política [43]. Los juegos serios son muy utilizados hoy en día en el aula, aunque son más aplicados a competencias específicas que a genéricas.
- *Herramientas para el análisis de los registros de aprendizaje (Learning analytics tools)*: herramientas que facilitan el análisis de los registros de los entornos de aprendizaje proporcionando un entorno para la visualización de los mismos, con diferentes informes y representaciones gráficas.

### 2.3.3 Categorización del estudio

Una vez revisados todos los artículos, se extrajeron unas características comunes a la tipología de los trabajos.

CATEGORÍA	TRABAJOS
Evaluación entre iguales y autoevaluación	13
Evaluación del docente	11
Herramientas de evaluación automática	5
Revisiones de la literatura	1

**Tabla 2.5:** Distribución de publicaciones por tratamiento del problema

De los trabajos seleccionados, son 5 los que proponen la evaluación automática de competencias genéricas. De éstos, sólo dos mencionan un enfoque como el que se propone en la introducción de este capítulo, es decir, aprovechando los registros de interacción de los estudiantes con LMS como indicadores del desempeño de las competencias genéricas. Encontramos trabajos que se apoyan en la tecnología para el tratamiento o evaluación de las competencias, pero recae en el usuario la evaluación, ya sea mediante autoevaluación o evaluación entre iguales, mediante evaluaciones realizadas directamente por el docente o mediante ambos enfoques. En concreto, de los 13 trabajos seleccionados que aplican evaluación entre iguales o autoevaluación, 4 aplican también evaluación del docente. En la tabla 2.5 se puede ver la distribución de las publicaciones. Además, nos encontramos con una revisión de la literatura sobre las competencias genéricas más evaluadas. Dicha revisión se utiliza para contrastar los datos sobre esas competencias con los obtenidos en este mapeado para responder a la primera pregunta de investigación.

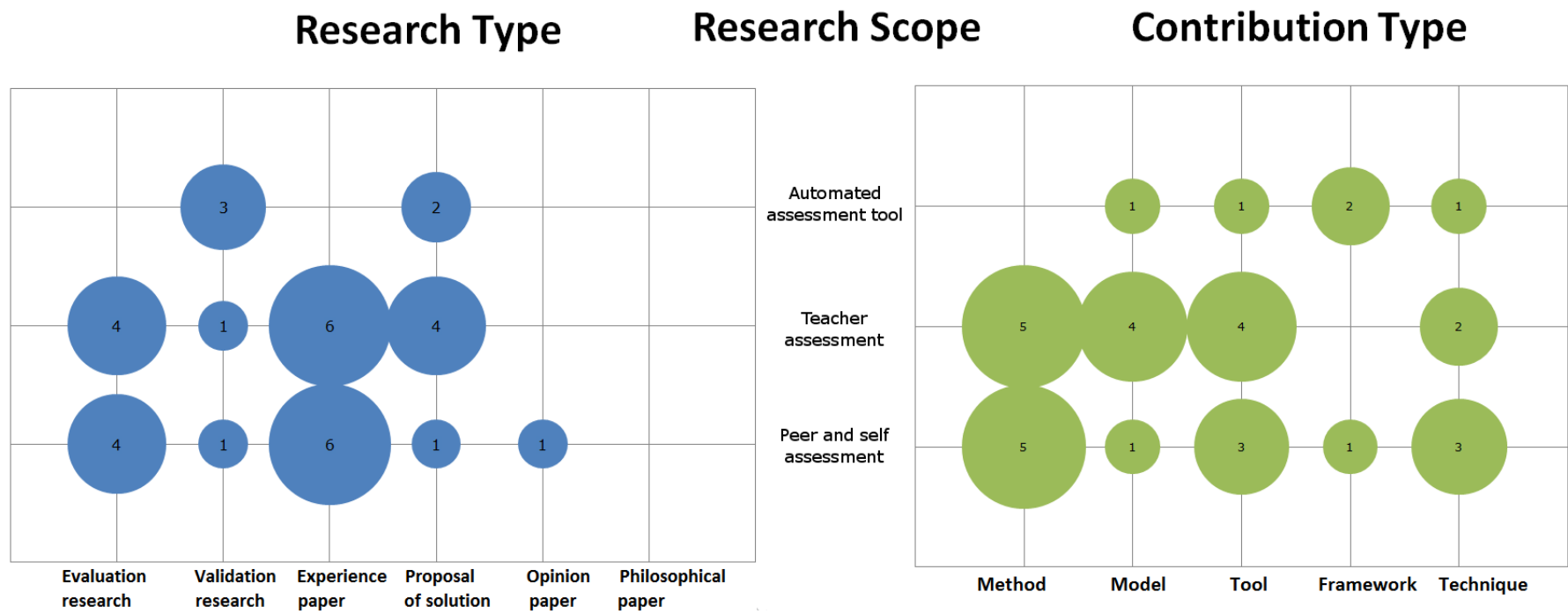
En la figura 2.2 se muestra la clasificación de los trabajos según su ámbito y su tipo (lado izquierdo), y según su ámbito y su contribución (lado derecho). La mayoría de los trabajos son propuestas (*Proposal of solution*), experiencias (*Experience papers*), validaciones (*Validation research*) y evaluaciones de la investigación (*Evaluation research*), mientras que casi no hay trabajos típicos de un tema de investigación con cierta madurez como los de opinión (*Opinion papers*) y los filosóficos (*Philosophical papers*).

El tipo de contribución está más distribuido. Las contribuciones del tipo proceso (*method*), modelo (*model*) y herramienta (*tool*) son las que se dan con más frecuencia: la primera con evaluación del docente (*teacher assessment*) y autoeva-

## 2. ESTADO DEL ARTE

---

luaciones o evaluaciones entre compañeros (*peer and self-assessment*), mientras que la segunda y la tercera se dan con más frecuencia con evaluaciones del docente. Cabe destacar que en esta figura si hay un trabajo que utiliza dos tipos de contribución, por ejemplo, evaluaciones del docente y evaluaciones entre iguales o autoevaluación, se contará una vez para cada tipo.



**Figura 2.2:** Ámbito de trabajos distribuidos según tipo de investigación y según tipo de contribución

## 2. ESTADO DEL ARTE

---

### 2.3.4 Esquema de clasificación

Todos los artículos seleccionados comparten dos características: primero, que evalúan competencias genéricas y, segundo, que la tecnología juega un papel en esta evaluación. Este papel puede ser para el desarrollo de las competencias, para la evaluación de las mismas o para ambos. Para esta tesis el aspecto que nos interesa es la evaluación, aunque hay trabajos en los que no se puede separar el cómo se han trabajado las competencias de cómo se han evaluado. El listado de trabajos se muestra en la tabla 2.12.

Desde el punto de vista de la evaluación hay dos tipos de trabajos:

- Evaluación asistida: en este grupo se engloban trabajos en los que la función de la herramienta informática para la evaluación de una o varias competencias genéricas de los estudiantes es la de dar soporte a la misma proporcionando el formato para introducir datos (notas, indicadores, respuestas a preguntas, etc.). Pero necesitan que alguien introduzca dichos datos. En base a ese “alguien” nos encontramos dos tipos de trabajo:
  - Autoevaluación o evaluación entre iguales.
  - Evaluación del docente.
- Evaluación semiautomática: se utiliza una herramienta informática que automatiza en parte la evaluación de los estudiantes. Puede ser que esta herramienta requiera una intervención inicial del docente para introducir datos o configurar la propia herramienta.

#### **Autoevaluación o evaluación entre iguales.**

La autoevaluación es un proceso en el que los estudiantes evalúan su propio trabajo, mientras que en el proceso de evaluación entre iguales un estudiante evalúa el trabajo de otro u otros estudiantes. Esta práctica se emplea por un lado para mejorar tanto el conocimiento en la materia del alumnado como sus habilidades metacognitivas, y por otro, para aliviar la carga de trabajo del profesorado. A menudo este tipo de evaluación se acompaña de algún tipo de rúbrica [67].

En algunos trabajos el proceso de evaluación de competencias genéricas se lleva a cabo después de haber trabajado los estudiantes en actividades de un entorno

virtual de aprendizaje mediante cuestionarios que completan los propios estudiantes o el docente. Los docentes confirman que el uso de herramientas web mejora la participación en la comunidad favoreciendo la colaboración y la construcción de conocimiento compartido [95]. Las herramientas que suelen utilizarse son wikis, foros, actividades y e-portfolio. Esta última aparece en varios trabajos, un e-portfolio (del inglés *electronic portfolio*), consiste en un conjunto de documentos, generalmente textos, archivos e imágenes, gestionados en un entorno web por un usuario. Los estudiantes trabajan con esta herramienta durante el curso y al final autoevalúan el desempeño de alguna competencia genérica [7].

Hay trabajos que implementan una metodología de *aprendizaje basado en problemas* (PBL, del inglés *Problem-Based Learning*) para desarrollar competencias específicas y genéricas en sus estudiantes y que después se evalúan mediante autoevaluación y evaluación entre iguales [69]. En [64] los docentes llevaron a cabo la evaluación del 90% de las competencias utilizando la herramienta de rúbricas *RubiStar*, mientras que los estudiantes mediante autoevaluación y evaluación entre iguales se encargaron del otro 10%.

En otro trabajos los estudiantes realizan una experiencia de *aprendizaje basado en equipos* (TBL, del inglés *Team-Based Learning*) utilizando algún tipo de herramienta colaborativa. En [44] se presenta un modelo que persigue el aprendizaje basado en equipos para la adquisición y evaluación de competencias genéricas en un contexto de e-learning. Los estudiantes trabajaban en grupo y evaluaban su desempeño en el *trabajo en equipo* mediante una rúbrica. La calificación se completó con un cuestionario.

Se encontraron varias experiencias que utilizan herramientas ligadas al ámbito empresarial (contabilidad, gestión de equipos, gestión de proyectos, etc.) que vienen acompañadas de rúbricas de autoevaluación. En [22] se utiliza la herramienta *Cycloid* para el desarrollo de competencias en la gestión de proyectos y posteriormente se llevan a cabo autoevaluaciones de los propios estudiantes para valorar la adquisición de dichas competencias. También se autoevalúan competencias *empresariales* en [1] mediante el uso de *Tricuspoid*.

Otra experiencia se lleva a cabo a partir de herramientas de videoconferencia como Skype o Hangouts. Los estudiantes realizan videoconferencias y después son

## 2. ESTADO DEL ARTE

---

autoevaluados o evaluados por sus compañeros en las competencias que deberían desempeñar en la actividad [70].

A continuación se van a enumerar las competencias que se han evaluado, los métodos que se han seguido y las técnicas y herramientas que se han utilizado para llevar a cabo estas evaluaciones.

### - Competencias evaluadas

Como se puede ver en la tabla 2.6, el *trabajo en equipo* (6 trabajos) y la *comunicación* (5 trabajos) son las competencias más evaluadas mediante evaluación entre iguales y autoevaluación. En el típico caso que encontramos en este grupo los estudiantes trabajan en grupo y evalúan su desempeño y el de sus compañeros en el *trabajo en equipo* mediante una rúbrica [44]. El uso de tecnologías de la información (*TIC*) y el *idioma* son las otras competencias que destacan dentro de este grupo con 3 trabajos cada una.

### - Métodos, técnicas e instrumentos de evaluación

En la tabla 2.7 se indican los métodos, técnicas e instrumentos de evaluación seguidas por cada uno de los trabajos que sigue este enfoque de autoevaluación y evaluación entre iguales. El método de evaluación más empleado es el sumativo (8 trabajos), mientras que las técnicas de observación sistemática y las pruebas escritas son los métodos más utilizados con 5 y 6 trabajos respectivamente, utilizando como herramientas cuestionarios y rúbricas. Podemos decir que el patrón clásico de este tipo de trabajos consiste en que los estudiantes se autoevalúan o evalúan a sus compañeros mediante rúbricas y cuestionarios una vez terminado el trabajo que debían realizar en el entorno virtual de aprendizaje.

### - Análisis

Aunque la autoevaluación y evaluación entre iguales son enfoques que pueden reducir el trabajo del docente, no siempre es así. En algunos trabajos anteriores, este enfoque a menudo sólo se utiliza de manera complementaria a algún otro tipo de evaluación [64, 94]. Además, se puede dar el caso que la autoevaluación no se ajuste del todo a la realidad del desempeño del estudiante. Por ejemplo, en [21]



## 2.3 Resultados

COMPETENCIA	CANT.	TRABAJOS
Análisis	2	[21, 64]
Aprendizaje permanente	1	[76]
Comunicación	5	[21, 69, 70, 76, 92]
Creatividad	1	[81]
Cultural	1	[76]
Emprendimiento	2	[1, 22]
Gestión de proyectos	1	[69]
Habilidades interpersonales	1	[69]
Investigación	1	[76]
Lengua extranjera	3	[70, 88, 94]
Liderazgo	1	[69]
Pensamiento crítico	1	[7]
Planificación y gestión del tiempo	1	[69]
Resolución de problemas	1	[76]
Responsabilidad	2	[21, 92]
TIC	3	[64, 70, 76]
Trabajo autónomo	1	[64]
Trabajo en equipo	6	[21, 44, 64, 69, 76, 81, 92]

**Tabla 2.6:** Competencias genéricas evaluadas mediante autoevaluación y evaluación entre iguales

## 2. ESTADO DEL ARTE

MÉTODOS DE EVALUACIÓN	TÉCNICAS DE EVALUACIÓN	INSTRUMENTOS DE EVALUACIÓN	TRABAJOS
Auténtica	Actividades sin determinar	-	[88]
	Observación sistemática	Rúbrica	[76]
Formativa	Pruebas escritas	Cuestionario	[1]
	Observación sistemática	Rúbrica	[7, 81]
Sumativa	Pruebas orales		[21, 64]
	Pruebas escritas	Cuestionario	[70]
			[22, 44, 69, 92, 94]

**Tabla 2.7:** Instrumentos de evaluación y métodos correspondientes a los trabajos de la autoevaluación y evaluación entre iguales

hay notables diferencias entre las calificaciones que se auto-asignan los estudiantes en algunas competencias y las calificaciones que le asignaron los docentes en esas mismas competencias. En ese trabajo se promovió la adquisición de competencias genéricas desde un punto de vista interdisciplinar y se diseñaron herramientas específicas para evaluar dichas habilidades. A la hora de evaluar, se realizaron tanto autoevaluaciones como evaluaciones del docente. En dicha experiencia se evaluaron cuatro competencias genéricas: *capacidad de análisis*, *habilidades de escritura*, *responsabilidad* y *capacidad de trabajo en equipo*. Cabe destacar discrepancias significativas entre las calificaciones que se auto-asignan los estudiantes en las dos primeras competencias. En la *capacidad de análisis* la discrepancia es de un 55,65%, mientras que en las *habilidades de escritura* de un 13,75%.

### Evaluación del docente

En esta sección se incluyen trabajos en los que la evaluación la realiza directamente el docente apoyándose en la tecnología. En su mayoría son trabajos en los que el docente ha de corregir cuestionarios con preguntas abiertas, en los que utiliza

rúbricas para evaluar el trabajo de sus estudiantes en las herramientas del entorno virtual de aprendizaje o en los que evalúa competencias genéricas a partir de las calificaciones de los estudiantes en las actividades que han realizado durante el curso.

En este tipo de trabajos se repiten experiencias ya vistas en la autoevaluación y evaluación entre iguales para la evaluación de competencias genéricas. Trabajos en los que la evaluación se realiza después de haber trabajado los estudiantes con herramientas del entorno virtual de aprendizaje [95], metodologías PBL [63] o experiencias basadas en videoconferencias [102].

En [105] se implementa un itinerario de aprendizaje incluyendo la evaluación de competencias genéricas. Un itinerario de aprendizaje es un mapa conceptual que nos guía en el proceso de aprendizaje. En este tipo de trabajos se implementan los itinerarios de aprendizaje definiendo matemáticamente tanto las fórmulas necesarias para evaluar cada bloque del curso a partir de sus actividades como su combinación para evaluar los objetivos del curso y las competencias genéricas.

### - Competencias evaluadas

Las competencia genérica más evaluada mediante la evaluación del docente es la *comunicación* oral y escrita (6 trabajos). En estos trabajos los docentes evalúan a los estudiantes mediante la corrección de documentos escritos y presentaciones orales. También le siguen de cerca los trabajos que evalúan las competencias de *trabajo en equipo* y de *resolución de problemas* (5 trabajos cada una). En la tabla 2.8 puede ver la relación de competencias evaluadas y los trabajos.

### - Métodos, Técnicas e instrumentos de evaluación

Dentro de la evaluación del docente el método de evaluación más seguido vuelve a ser el sumativo (5 trabajos), seguido de cerca por el formativo (4 trabajos). La técnica más utilizada es la observación sistemática (6 trabajos) y la herramienta más utilizada, la rúbrica (4 trabajos). Puede verse el listado completo de métodos, herramientas y técnicas en la tabla 2.9.

## 2. ESTADO DEL ARTE

COMPETENCIA	CANT.	TRABAJOS
Análisis	1	[8]
Aprendizaje permanente	1	[84]
Comunicación	6	[17, 63, 68, 84, 89, 105]
Emprendimiento	2	[84, 102]
Pensamiento crítico	2	[8, 63]
Planificación y gestión del tiempo	1	[63]
Resolución de problemas	5	[8, 17, 68, 89, 101]
Trabajo autónomo	1	[64]
Trabajo en equipo	5	[17, 63, 68, 84, 89]

**Tabla 2.8:** Competencias evaluadas directamente por el docente

MÉTODOS DE EVALUACIÓN	TÉCNICAS DE EVALUACIÓN	INSTRUMENTOS DE EVALUACIÓN	TRABAJOS
Auténtica	Actividades sin determinar	-	[95]
Con referencia al criterio			[93]
Formativa	Pruebas orales	Entrevista	[102]
	Observación sistemática	Herramienta de seguimiento	[17, 63]
		Rúbrica	[89]
Sumativa	Actividades sin determinar	-	[8, 68, 84]
	Pruebas escritas	Cuestionario	[105]
			[101]

**Tabla 2.9:** Instrumentos de evaluación y métodos correspondientes a los trabajos en los que la evaluación es realizada por los docentes

### - Análisis

La escalabilidad es el problema más mencionado por los autores en los trabajos recopilados. En [93] se diseña *Hiperion*, un sistema de recomendación que ayuda a diseñar actividades adaptadas a cada estudiante para mejorar sus competencias. En su estudio de caso los docentes evaluaban las competencias de los estudiantes manualmente y después aplicaban Hiperion. La principal desventaja de la herramienta es el tiempo que el docente ha de dedicar para asignar los diferentes logros y el peso de cada nota para cada competencia en las actividades. En línea con los problemas de escalabilidad anteriores nos encontramos con el trabajo mostrado en [63]. En él se utiliza una metodología PBL, en la que se realiza una evaluación individualizada de cada estudiante y de cada grupo de estudiantes. El autor considera también que el esfuerzo necesario y carga de trabajo para cada docente es un poco mayor al habitual. Lo mismo ocurre en [17], trabajo en el que los docentes concluyeron que el esfuerzo que realizaron fue excesivo a pesar de los buenos resultados obtenidos y descartaron el uso del portfolio para próximas experiencias ya que les supone una gran carga de trabajo sobre todo en el tramo final del curso.

### Evaluación semiautomática

Las herramientas de evaluación semiautomática son herramientas informáticas que ayudan al docente en el proceso de la evaluación automatizando dicho proceso y proporcionando una calificación o indicador para cada estudiante susceptible de ser aplicado a la evaluación de su desempeño en una o varias competencias genéricas. Estas herramientas requieren una intervención inicial del docente para introducir datos o configurar la herramienta.

En [5] se propuso un modelo formal para asignar trabajadores a proyectos software. Para definir el modelo se siguió un método Delphi, donde un grupo de expertos definieron criterios para la evaluación de habilidades de trabajo en equipo y definieron un test psicológico.

Hay trabajos en los que se emplea una metodología de aprendizaje basado en juegos (GBL, del inglés *Game-Based Learning*) [15, 53]. En ellos se utilizan juegos digitales con el fin de apoyar y mejorar la enseñanza, el aprendizaje y la evaluación. Se puede establecer como una actividad práctica donde los jugadores, a

## 2. ESTADO DEL ARTE

---

COMPETENCIA	CANT.	TRABAJOS
Análisis	1	[5]
Aprendizaje permanente	1	[5]
Comunicación	3	[5, 15, 85]
Creatividad	1	[5]
Emprendimiento	1	[53]
Gestión de proyectos	1	[5]
Habilidades interpersonales	2	[5, 85]
Investigación	1	[5]
Liderazgo	1	[5]
Pensamiento crítico	1	[5]
Planificación y gestión del tiempo	1	[5]
Resolución de problemas	1	[53]
Responsabilidad	1	[5]
Toma de decisiones	1	[5]
Trabajo autónomo	1	[5]
Trabajo en equipo	2	[5, 45]

**Tabla 2.10:** Competencias evaluadas de forma semiautomática

medida que van avanzado en las dinámicas del juego, deben evidenciar unas habilidades, conocimientos y competencias que muestran el alcance de los objetivos de aprendizaje [23].

Y finalmente se han encontrado trabajos que se basan en técnicas de *learning analytics* para obtener indicadores con los que evaluar la competencia de trabajo en equipo a partir de la actividad de los estudiantes en diferentes entornos virtuales de aprendizaje [45, 85].

### - Competencias evaluadas

En la tabla 2.10 se muestran las competencias evaluadas y los trabajos en que se evalúan. Aunque no son muchos trabajos los que encontramos en este grupo, cabe destacar que vuelve a ser la competencia de la *comunicación* la más evaluada.

MÉTODOS DE EVALUACIÓN	TÉCNICAS DE EVALUACIÓN	INSTRUMENTOS DE EVALUACIÓN	TRABAJS
Auténtico	Indicadores basados en logros	Juegos serios	[15]
Formativo			Indicadores de trabajo en entornos virtuales de aprendizaje
	Sumativo	Pruebas escritas	Test automático

**Tabla 2.11:** Instrumentos de evaluación y métodos correspondientes a los trabajos de evaluación semiautomática

### - Métodos, Técnicas e instrumentos de evaluación

El método de evaluación formativo es el más empleado (3 trabajos). Al haber pocos trabajos no hay ninguna tendencia que destacar con respecto a las técnicas y las herramientas. Se puede ver cada método junto con las técnicas y herramientas empleadas en la tabla 2.11.

### - Análisis

La creación de los tests de personalidad no está al alcance de todos los docentes. Como hemos visto antes en el trabajo presentado en [5], fue necesario llevar a cabo un modelo Delphi para la definición de los tests.

En [53] se utilizan los juegos serios para el desarrollo de las competencias de *emprendimiento y solución de problemas*. Se definieron una serie de indicadores como medida del desempeño en las competencias que permiten al estudiante conocer su nivel de adquisición de las mismas. En [15] también se utilizan los juegos serios para el desarrollo y evaluación de competencias genéricas. Se basa en un modelo donde para cada competencia se identifican subcompetencias más específicas, lo que facilita el proceso de definición de indicadores. Los juegos serios suelen utilizarse con un propósito específico, generalmente relacionado con competencias específicas. Hay muchos trabajos sobre juegos serios en la literatura pero sólo dos que evalúen competencias genéricas.

## 2. ESTADO DEL ARTE

---

El primer trabajo en el que se usan los registros de actividad de los entornos de aprendizaje utiliza LACAMOLC, una plataforma web que aporta información visual e informes con indicadores de competencias genéricas de los estudiantes a partir de los registros de actividad [85]. LACAMOLC está implementado sobre Pentaho. Pentaho es una herramienta de análisis de negocio que recoge datos de las bases de datos de los diferentes orígenes, y que mapeará estos datos con los indicadores de las competencias genéricas. Las competencias que se evalúan y los indicadores que se utilizan son:

- *Gestión del tiempo*. Indicador: cumplimiento de la planificación, formulado en base al número de acciones que se han hecho a tiempo con respecto a la planificación fijada en una hoja de cálculo de Google Docs.
- *Comunicación interpersonal*. Indicador: escuchar a los demás, formulado en base al número de veces que los estudiantes accedieron a las discusiones del foro del LMS Moodle.
- *Comunicación interpersonal*. Indicador: expresar sus ideas, formulado en base a las intervenciones en el foro de Moodle y comentarios en un documento Google Docs.
- *Habilidades de escritura*. Indicador: expresar sus ideas, formulado en base al número de veces que cada estudiante intervino en los foros de Moodle dividido por el número de veces que accedió.
- *Habilidades de escritura*. Indicador: expresar sus ideas con claridad y precisión, formulado en base al número total de palabras que cada estudiante escribió dividido por el número de veces que intervino.
- *Trabajo en equipo*. Indicador: participación activa, formulado en base al tiempo total que duran las discusiones divididos por el número de sesiones.
- *Pensamiento analítico*. Indicador: respaldo de ideas de otros, formulado en base al número total de mensajes que un estudiante ha leído dividido entre el número de veces que intervino.



- *Pensamiento analítico*. Indicador: identificación de errores o falta de coherencia en sus propias ideas, formulado en base al número de veces que un estudiante releyó sus comentarios.

En el segundo trabajo que utiliza los registros de aprendizaje se utiliza el método CTMTC (*Comprehensive Training Model of the Teamwork Competence*), que integra herramientas que están presentes en diferentes entornos virtuales de aprendizaje y que facilita el registro de la interacción de los estudiantes [45]. El registro de esta interacción en el foro era una tarea tediosa para ser realizada a mano, por lo que implementaron el *LA system*, un servicio web en Moodle, que permite que la información sea accesible a través de internet con mensajes basados en XML. Esta información después es consumida por un cliente que visualiza los datos. Los indicadores utilizados para evaluar el trabajo en equipo son:

- Mensajes escritos en foro: interacción estudiante-estudiante activo.
- Mensajes leídos en el foro: interacciones estudiante-estudiante pasivo.

Estos dos trabajos proporcionan indicadores de los entornos virtuales de aprendizaje aplicables a varias competencias genéricas. El inconveniente que presentan ambas aproximaciones es que los indicadores son fijos, es decir, el docente no tendría la opción de combinar o buscar otros indicadores en el entorno virtual aparte de los que proporcionan los métodos presentados. Por ejemplo, si en un LMS se utilizan los foros con unas reglas de uso diferentes quizás habría que modificar los indicadores o combinarlos de otra manera para que fueran útiles para el docente. Sería deseable la opción de que el docente pudiera diseñar sus propias evaluaciones a partir de los indicadores que proporciona el entorno virtual de aprendizaje.

REF	TÍTULO	TIPO DE INVESTIGACIÓN	TIPO DE CONTRIBUCIÓN	ÁMBITO DE LA INVESTIGACIÓN
[105]	A Fine-Grained Outcome-Based Learning Path Model	proposal of solution	model	Teacher assessment
[85]	A web platform for the assessment of competences in Mobile Learning Contexts	validation research	framework	Automated assessment tool
[68]	Acquired Skills With The Implementation Of New Evaluation Methods At University Rey Juan Carlos	experience paper	model	Teacher assessment
[63]	Active learning through problem based learning methodology in engineering education	experience paper	method	Teacher assessment
[17]	Adapting teaching and assessment strategies to enhance competence-based learning in the framework of the european convergence process	proposal of solution	method	Teacher assessment
[8]	Appraisal of Course Learning Outcomes using Rasch Measurement: A Case Study in Information Technology Education	proposal of solution	model	Teacher assessment
[94]	Assessment of competences in designing online preparatory materials for the Cambridge First Certificate in English examination	evaluation research	technique	Peer and self-assessment / Teacher assessment
[101]	Assessment of problem solving in computing studies	experience paper	method	Teacher assessment
[1]	Competence Assessment in Higher Education: A Dynamic Approach	proposal of solution	tool	Peer and self-assessment
[96]	Criteria and standards of generic competences at bachelor degree level: A review study	evaluation research	method	Review study

[102]	Developing entrepreneurial accounting and finance competency using the ELLEIEC Virtual Centre for Enterprise	experience paper	tool	Teacher assessment
[44]	e-Learning and Team-based Learning. Practical Experience in Virtual Teams	experience paper	framework	Peer and self-assessment
[84]	Engineering Students Performance Evaluation of Generic Skills Measurement: ESPEGS Model	validation research	model	Teacher assessment
[89]	e-Portfolio: A tool to assess university students' skills	evaluation research	tool	Teacher assessment
[5]	Formal model for assigning human resources to teams in software projects	validation research	model	Automated assessment tool
[15]	From Behavioral Indicators to Contextualized Competence Assessment	proposal of solution	framework	Automated assessment tool
[76]	Graduate attributes as a focus for institution-wide curriculum renewal: innovations and challenges	opinion paper	model	Peer and self-assessment
[93]	Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences	proposal of solution	tool	Teacher assessment
[22]	International creative tension study of university students in South Korea and Finland	evaluation research	tool	Peer and self-assessment
[81]	Measuring collaboration and creativity skills through rubrics: Experience from UTPL collaborative social networks course	evaluation research	method	Peer and self-assessment / Teacher assessment
[64]	Problem Based Learning Implementation In The Degree Of Human Nutrition And Dietetics	experience paper	technique	Peer and self-assessment / Teacher assessment

[7]	Promoting reflection on science, technology, and society among engineering students through an EAP online learning environment	evaluation research	tool	Peer and self-assessment
[70]	Self-video recording for the integration and assessment of generic competencies	experience paper	technique	Peer and self-assessment
[53]	Serious Games for the Development of Employment Oriented Competences	validation research	tool	Automated assessment tool
[92]	Soft Skills: A Comparative Analysis Between Online and Classroom Teaching	experience paper	method	Peer and self-assessment
[95]	Sustaining Teacher's Professional Development and Training through Web-Based Communities of Practice	evaluation research	tool	Teacher assessment
[88]	Teaching And Learning Through Projects Using The ICT: Practice Of The English Writing Through Business Documents	experience paper	method	Peer and self-assessment
[69]	Teamwork competence and academic motivation in computer science engineering studies	validation research	method	Peer and self-assessment
[21]	The promotion and assessment of generic skills from interdisciplinary teaching teams	experience paper	method	Peer and self-assessment / Teacher assessment
[45]	Using Learning Analytics to improve teamwork assessment	proposal of solution	technique	Automated assessment tool

**Tabla 2.12:** Distribución de publicaciones por tratamiento del problema

### 2.4 Respuestas a las preguntas de investigación

En base al estudio las respuestas a las preguntas de investigación son las siguientes:

#### **Q1. ¿Qué competencias se han evaluado de forma automática o asistida por ordenador a partir de la actividad de los estudiantes en los entornos virtuales de aprendizaje?**

En esta selección de trabajos se han encontrado evaluaciones para todas las competencias que fueron definidas en la tabla 2.4. Aunque las más evaluadas son la *comunicación* (14 trabajos), *trabajo en equipo* (13 trabajos) y *resolución de problemas* (7 trabajos). El resumen completo de número de trabajos por competencia puede verse en la tabla 2.13.

Estos datos se pueden contrastar con los de la revisión de la literatura mostrada en [96], donde se analizan las competencias genéricas más frecuentemente evaluadas. En este caso repite primera posición la competencia de la *comunicación*. La competencia de *resolución de problemas* ocupa la cuarta posición, mientras que la de *trabajo en equipo* pasa a sexto lugar. En esta revisión de la literatura se hallaron más artículos para todas las competencias, ya que no se descartaron por no ser procesos soportados tecnológicamente. El hecho de que haya más trabajos en proporción para la competencia de *trabajo en equipo* en nuestra selección de trabajos se puede deber a que los entornos virtuales de aprendizaje modernos suelen favorecer los procesos de cooperación [54].

#### **Q2. ¿Qué métodos se utilizan para evaluar competencias genéricas en entornos virtuales de aprendizaje?**

Se han encontrado dos grupos de trabajo, por un lado, aquellos en los que el entorno virtual de aprendizaje asiste al usuario en la evaluación, y por el otro, aquellos en los que el entorno virtual de aprendizaje realiza la evaluación automáticamente. Dentro del primer grupo, estos trabajos se dividen en dos subgrupos dependiendo de quién realice la evaluación: evaluación entre iguales o autoevaluación (cuando la evaluación la realizan los estudiantes) y evaluación del docente.

Estos grupos se han separado en la categorización final debido al elevado número de trabajos que hay para cada uno de ellos. Además, las competencias que se

## 2. ESTADO DEL ARTE

---

COMPETENCIA	TRABAJOS
Análisis	4
Aprendizaje permanente	3
Comunicación	14
Creatividad	3
Cultural	1
Emprendimiento	5
Gestión de proyectos	2
Habilidades interpersonales	3
Investigación	2
Lengua extranjera	3
Liderazgo	2
Pensamiento crítico	4
Planificación y gestión del tiempo	3
Resolución de problemas	7
Responsabilidad	3
TIC	3
Toma de decisiones	1
Trabajo autónomo	2
Trabajo en equipo	13

**Tabla 2.13:** Número de trabajos que evalúan cada competencia genérica

## 2.4 Respuestas a las preguntas de investigación

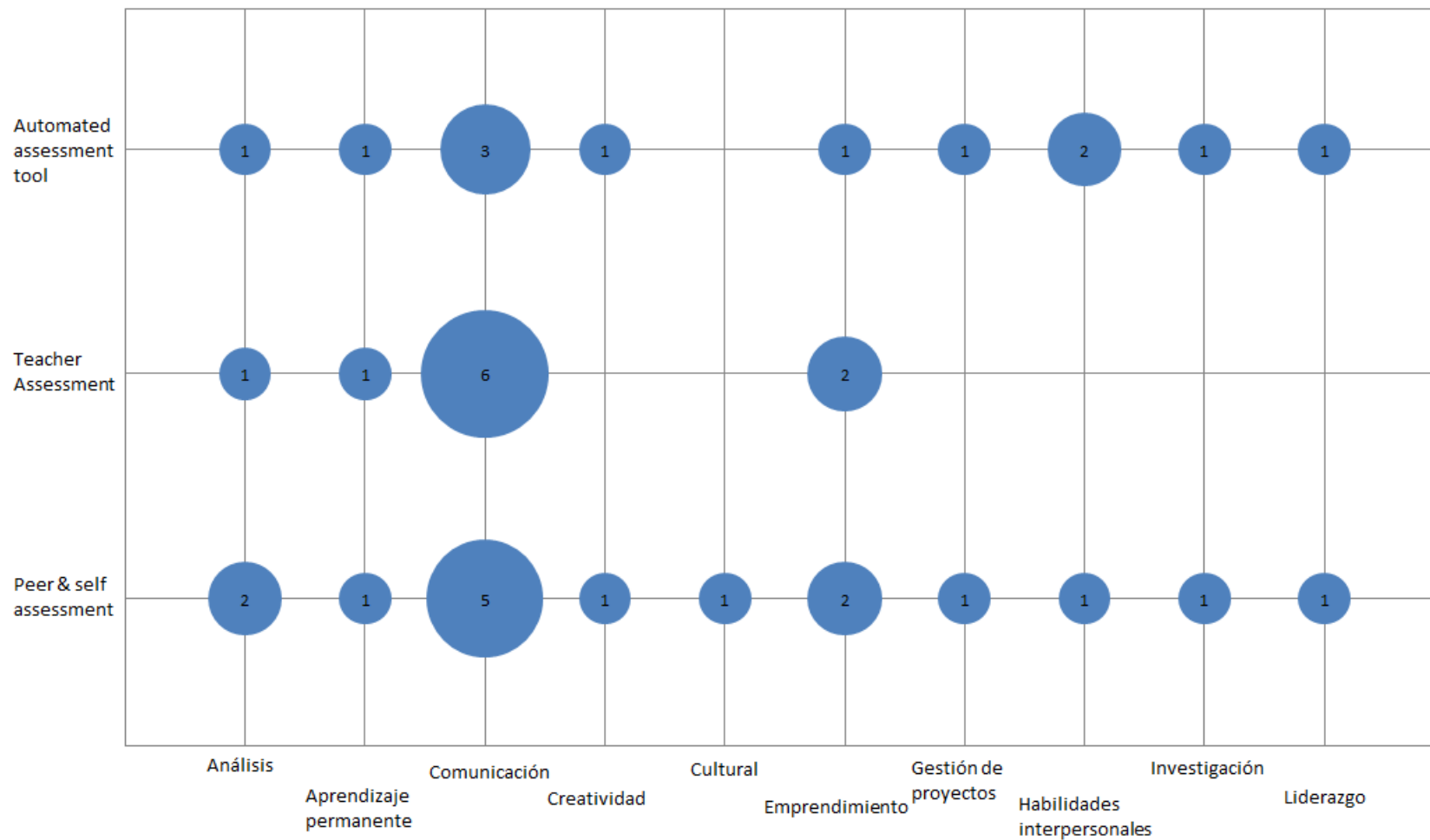
---

evalúan no son las mismas con un enfoque u otro. Mientras que ambos enfoques tienen un número elevado de trabajos que evalúan las competencias de *comunicación y trabajo en equipo*, la competencia de *resolución de problema* tiene una presencia mucho más significativa en trabajos en los que evalúa el docente (5 trabajos) con respecto a trabajos que evalúan los estudiantes (1 trabajo).

La herramienta que más se utiliza tanto en la evaluación del docente como en la evaluación entre iguales o autoevaluación es la rúbrica electrónica. El problema que encontramos es estos trabajos que utilizan rúbricas es que si el docente se encarga de la evaluación, la carga de trabajo de éste aumenta [63]. Sin embargo, si se delega en la autoevaluación o evaluación entre iguales pueden aparecer discrepancias entre las calificaciones que se auto-asignan los estudiantes y las que reciben por otros medios [21].

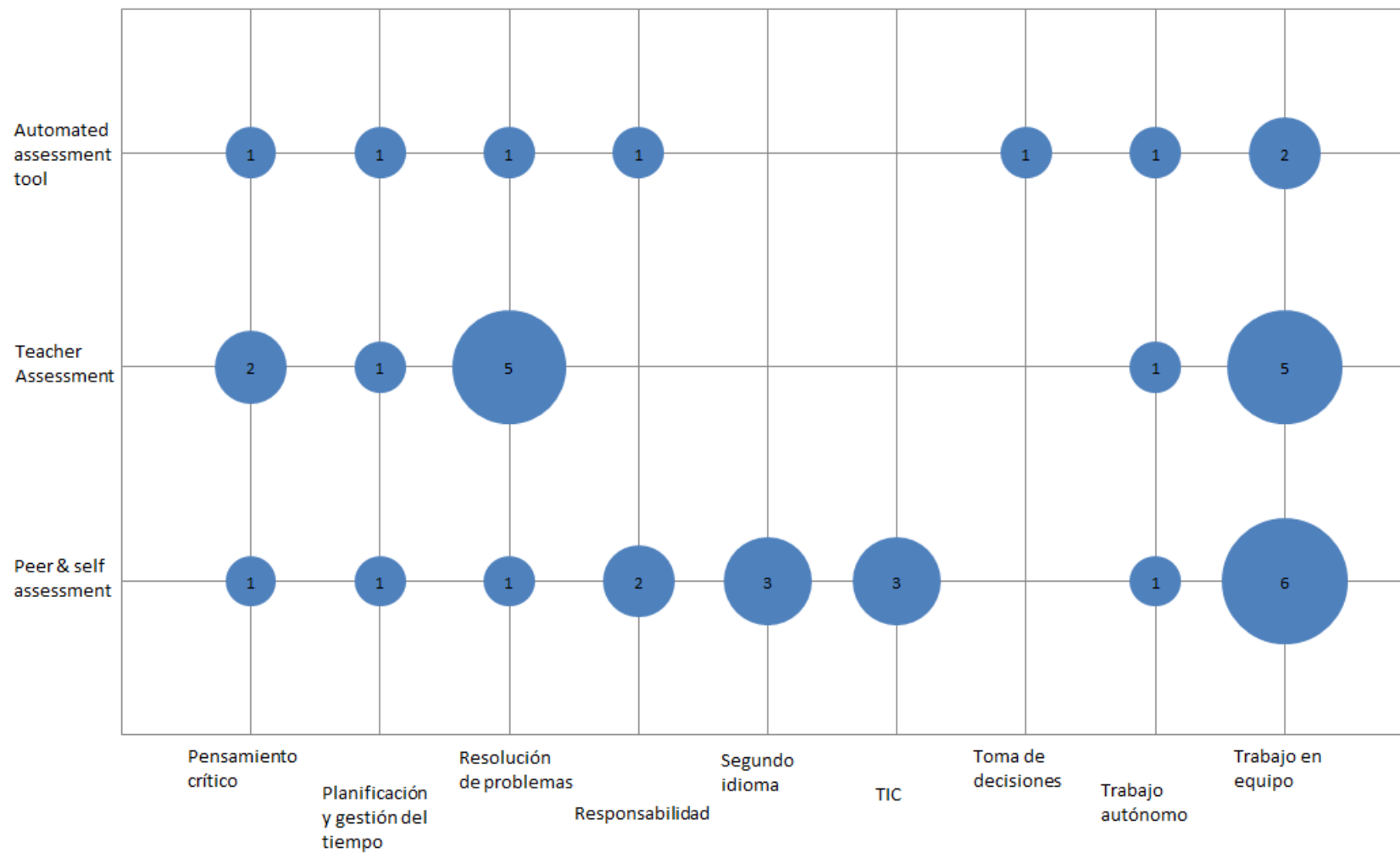
Dentro de la evaluación semiautomática se seleccionaron 5 trabajos, entre los que se encuentran algunos basados en juegos serios [15, 36] y otros basados en el análisis de los procesos de aprendizaje [45, 85].

Pueden verse las competencias genéricas evaluadas con cada método en las figuras 2.3 y 2.4.



**Figura 2.3:** Competencias genéricas evaluadas con cada método (1 de 2)





**Figura 2.4:** Competencias genéricas evaluadas con cada método (2 de 2)

## 2. ESTADO DEL ARTE

---

### Q3. ¿Qué técnicas se utilizan para evaluar competencias genéricas a partir de los registros de actividad de un entorno virtual de aprendizaje?

Dentro de los trabajos que realizan una evaluación semiautomática hay dos trabajos que utilizan un método de evaluación formativo a partir de técnicas para la extracción de indicadores de trabajo en el entorno virtual de aprendizaje utilizando herramientas para el análisis de los registros de aprendizaje [45, 85].

En [85] se utilizan procesos ETL (Extract, Transform and Load) mediante el uso de la herramienta de análisis Pentaho para obtener los registros de actividad de diversos KLT (Knowledge and Learning Technologies) como Moodle, Google Apps for Education, Mediawiki y Edmodo.

En [45] se utilizan procesos de minería de datos (*data mining*) mediante el desarrollo de un servicio web en Moodle que devuelve información almacenada en los registros en la plataforma de aprendizaje.

A continuación se muestra un resumen para estos trabajos indicando las técnicas empleadas, los métodos aplicados y las competencias evaluadas (tabla 2.14).

Competencias	Métodos	Técnicas	Herramientas
	Formativo	Indicadores de trabajo en entornos virtuales de aprendizaje	Herramienta para el análisis de los registros de aprendizaje
<i>Análisis</i>	[85]		
<i>Comunicación</i>			
<i>Habilidades interpersonales</i>			
<i>Planificación</i>			
<i>Trabajo en equipo</i>	[45]		

**Tabla 2.14:** Competencias evaluadas, métodos aplicados y técnicas seguidas para evaluar competencias a partir de los registros de actividad de los entornos de aprendizaje

### 2.5 Conclusiones

Las competencias genéricas son hoy en día una pieza fundamental en las planificaciones de la enseñanza a todos los niveles académicos. Como consecuencia de esto, son numerosos los trabajos y los proyectos que se han puesto en marcha en los últimos años para fomentar el desarrollo de las mismas en los estudiantes. Tanto durante el desarrollo como a la finalización de estos trabajos y proyectos, es deseable una evaluación del nivel de adquisición de estas competencias en los estudiantes.

En la literatura hemos encontrado diferentes problemas a la hora de afrontar esta evaluación. Por un lado problemas de objetividad, ya que los criterios que para un docente son válidos para la evaluación de una competencia genérica pueden no ser válidos para otro. Y por otro lado, problemas de escalabilidad. Si ya en muchos casos la carga de trabajo del profesorado para poder alcanzar los objetivos del curso es elevada, aún más lo será si éstos tienen además que generar y evaluar nuevas actividades para medir el nivel de desempeño de sus estudiantes en competencias genéricas. Es más, este problema de escalabilidad se acrecienta en el contexto de los LMS, donde los cursos en ocasiones contienen un número muy elevado de estudiantes (por ejemplo los cursos de tipo MOOC).

Para conocer el estado del arte en la evaluación de competencias genéricas mediante el uso de la tecnología se ha realizado un estudio de la literatura en forma de SMS. Tras esta revisión fueron 30 los trabajos seleccionados y a partir de ellos se ha dado respuesta a las preguntas de investigación inicialmente planteadas.

En muchos de estos trabajos es el docente quien realiza la evaluación apoyándose en diferentes herramientas, pero estos trabajos suelen presentar problemas de escalabilidad. También tenemos trabajos en los que se trata de minimizar esta carga de trabajo del docente combinando o sustituyendo la evaluación del docente con evaluaciones entre iguales o autoevaluaciones. Esto evita en parte el problema de la carga de trabajo, pero nos encontramos con problemas de objetividad en algunas evaluaciones de los estudiantes. Como consecuencia, el profesorado tiene que revisar las evaluaciones de sus estudiantes, por lo que se puede volver a encontrar con problemas de carga de trabajo.

Otros trabajos consisten en la implementación de alguna herramienta online de cuestionarios de personalidad, normalmente diseñados para revelar aspectos del

## 2. ESTADO DEL ARTE

---

carácter o mecanismos psicológicos de un individuo. El inconveniente es que realizar un cuestionario de este tipo no está al alcance de cualquier docente, sino que debe tener formación en psicología, especialmente para diagnosticar o valorar los resultados. Y aunque se trate de automatizar el procedimiento, en ocasiones estos tests están formados con preguntas cerradas y abiertas, y en el caso de estas últimas, volvemos a encontrarnos con problemas de escalabilidad.

Por último hemos encontrado un conjunto de trabajos que automatizan el proceso de evaluación de competencias genéricas. Por un lado encontramos juegos serios, juegos que emulan un caso real profesional y en base al modo de actuar del estudiante obtendrán una puntuación que servirá como medida del desempeño en ciertas competencias. Estos juegos suelen estar muy enfocados a ciertas competencias no genéricas y su implementación es costosa. Además, el proceso de extracción de calificaciones no está automatizado ni integrado con las herramientas de evaluación del docente, por lo que es necesario un procedimiento manual para capturarlas.

Por otro lado nos encontramos con el tipo de trabajo que cubren en mayor medida los objetivos de nuestra investigación: trabajos que evalúan competencias genéricas a partir de indicadores obtenidos de los registros de los entornos de aprendizaje. En estos trabajos las herramientas proporcionan unos indicadores fijos sobre actividades concretas, pero no dan al docente la opción de decidir fórmulas ni crear sus propias evaluaciones, así que lo que proporciona la herramienta podría ser válido para un docente pero no serlo para otro. Si tuviéramos un método que nos permitiera diseñar evaluaciones se podrían probar diferentes fórmulas para obtener diferentes indicadores del desempeño de los estudiantes en el entorno virtual de aprendizaje hasta encontrar aquellos que se adapten a lo que cada docente considere válido para la evaluación de competencias.

# Resumen de problemas encontrados

En el capítulo anterior se ha realizado una revisión de la literatura para responder a diversas cuestiones sobre la evaluación de competencias genéricas mediante el uso de métodos y herramientas informáticas, así como de cuáles son esos métodos y si usan los registros de actividad de los entornos virtuales de aprendizaje. A raíz de esta revisión han aflorado varios problemas para cada tipo de evaluación encontrado que son resumidos a continuación. También se indican en este capítulo los requisitos que deberá tener el método a desarrollar a tenor de los problemas anteriores.

## 3.1 Problemas encontrados

La explicación de los problemas encontrados se organiza según los tipos de evaluación de los trabajos (asistida y semiautomática). Para cada uno de ellos se resumirán los problemas encontrados.

### **3. RESUMEN DE PROBLEMAS ENCONTRADOS**

---

#### **3.1.1 Evaluación asistida**

En este tipo de evaluación la herramienta informática da soporte al usuario para que éste lleve a cabo la evaluación. Para ello le proporciona el formato para introducir datos (notas, indicadores, respuestas a preguntas, etc.). Estos datos podrán ser introducidos por los alumnos (autoevaluación o evaluación entre iguales) o por el docente (evaluación del docente).

#### **Autoevaluación o evaluación entre iguales**

La autoevaluación es un proceso en el que los estudiantes evalúan su propio trabajo, mientras que en el proceso de evaluación entre iguales un estudiante evalúa el trabajo de otros estudiantes. La mayor virtud de este tipo de evaluación es que mejora tanto el conocimiento en la materia del alumnado como sus habilidades metacognitivas. Otra ventaja de esta evaluación es que ahorra parte del trabajo del docente.

A continuación se van a describir los problemas encontrados por los autores a la hora de realizar una evaluación de una o varias competencias genéricas con un enfoque de autoevaluación o evaluación entre iguales.

#### **Subjetividad**

En ocasiones las calificaciones que asignan los estudiantes no se ajustan a la que recibirían del docente, habiendo diferencias notables entre las calificaciones que asignan diferentes estudiantes o el docente a un mismo trabajo. Este tipo de evaluaciones suelen ir acompañados de rúbricas para guiar el proceso de evaluación [67]. Sin embargo, no todas las competencias a evaluar y los aspectos derivados a tener en cuenta pueden ser recogidos en una rúbrica. Si a esto unimos la falta de madurez que pueden tener los estudiantes o una interpretación diferente que estos realicen de los criterios de evaluación, las diferencias entre la calificación que el docente daría a un trabajo y la que darían sus estudiantes podrían ser importantes [21].

#### Escalabilidad

Se dice que un proceso de evaluación sufre problemas de escalabilidad cuando el número de trabajos a evaluar crece y el evaluador no es capaz de abarcar este crecimiento. A priori podría pensarse que la autoevaluación o evaluación entre iguales son estrategias que ahorran trabajo al docente, ya que son los estudiantes los que se encargan de la evaluación. Sin embargo, más que ahorrar trabajo, podría decirse que el docente adopta otro papel, ya que de evaluar trabajos pasa a moderar o revisar las evaluaciones de los estudiantes [20]. Además, el tiempo que debiera dedicar el docente a revisar las evaluaciones de los trabajos puede ser en muchos casos superior al tiempo que dedicaría a evaluar los trabajos.

Por ejemplo, supongamos que un docente tiene un número dado de estudiantes en clase ( $n$ ) y el tiempo de evaluación que le requiere el trabajo realizado por cada estudiante es  $t_{eval}$ . Evaluar el trabajo de estos estudiantes le tomará un tiempo de  $T = n * t_{eval}$ . Aplicando el principio de invarianza multiplicativa esta es una actividad desde un punto de vista temporal de orden  $n$  ( $O(n)$ ).

Si optamos por una estrategia de evaluación entre iguales en la que cada estudiante evalúa a un número determinado de compañeros ( $k$ ) y el docente considera revisar las evaluaciones realizadas por sus estudiantes, teniendo en cuenta que la revisión de cada evaluación le requiere un tiempo  $t_{rev}$ , este proceso tendría un coste de  $T = k * n * t_{rev}$ .

Dependiendo del número de evaluaciones que tenga que realizar cada estudiante nos encontraremos en diversos casos:

- Mejor caso: cada estudiante sólo evalúa a un único compañero ( $k = 1$ ). En este caso  $T = n * t_{rev}$ . Por el principio de invarianza multiplicativa este proceso sería también de ( $O(n)$ ).
- Peor caso: cada estudiante evalúa a todos sus compañeros ( $k = n - 1$ ). En este caso  $T = n * (n - 1) * t_{rev} = n^2 * t_{rev} - n * t_{rev}$ . Aplicando el principio de invarianza multiplicativa y la del máximo de dos órdenes queda que este caso es de orden cuadrático ( $O(n^2)$ ).

Es difícil establecer un caso promedio o general, pero si un docente fijase que cada estudiante tuviese que evaluar a la mitad de sus compañeros ( $n/2$ ) o a un tercio

### 3. RESUMEN DE PROBLEMAS ENCONTRADOS

---

$(n/3)$ , el orden del proceso, aplicando el principio de invarianza multiplicativa, seguiría siendo cuadrático ( $O(n^2)$ ).

Por tanto, en el mejor caso, ambas estrategias no difieren en su implementación en más de alguna constante multiplicativa, siendo ambas estrategias de orden  $n$  ( $O(n)$ ). Pero en el caso general, el orden del proceso de revisión de las autoevaluaciones realizadas por los estudiantes es cuadrático ( $O(n^2)$ ), mayor que el orden de tiempo que conlleva la evaluación del docente de los trabajos ( $O(n)$ ). Por lo que puede afirmarse que este enfoque no resuelve el problema de la escalabilidad.

En algunos de los trabajos revisados, puede verse como tras la evaluación realizada por los estudiantes, los docentes han de revisar dichas evaluaciones [64, 94]. Los autores indican la sobrecarga de trabajo que ello les supuso, quedando por tanto ilustrado lo mencionado en el análisis anterior.

#### **Evaluación del docente**

En esta sección se recogen problemas asociados a los trabajos en los que la evaluación la realiza directamente el docente apoyándose en la tecnología.

#### **Escalabilidad**

La escalabilidad vuelve a ser el problema más mencionado en este tipo de evaluación. Los autores de los trabajos comparten experiencias en las que evaluaron competencias genéricas, pero indican que la carga de trabajo les resultó excesiva [63, 93] e incluso uno de ellos descarta repetir la experiencia en los siguientes cursos debido al sobreesfuerzo que le supuso [17]. Podemos deducir, por tanto, que si en ocasiones los docentes tienen tiempo ajustado para lograr los objetivos curriculares específicos de la materia que tienen que impartir y las evaluaciones que tienen que realizar, más difícil será lograr dichos objetivos si tienen que diseñar tareas adicionales para que los alumnos desempeñen competencias genéricas y después evaluarlas.



### 3.1.2 Evaluación semiautomática

En esta sección se recogen problemas asociados a trabajos que utilizan herramientas de evaluación semiautomática. Estas herramientas ayudan al docente en la evaluación proporcionándole de manera automática calificaciones o indicadores que éste puede usar para la evaluación de sus estudiantes. La intervención del docente en estas herramientas consiste en configuraciones iniciales de parámetros del curso o la evaluación.

La automatización de este proceso de evaluación solventa el problema de escalabilidad mencionado en los enfoques anteriores. Pero se encuentran otros problemas que serán descritos a continuación.

#### **Propósito específico**

Los recursos con los que cuentan los docentes están enfocados o fueron creados para un propósito específico. Dentro de los trabajos seleccionados esto ocurre principalmente con los juegos serios, que son diseñados para un propósito principal distinto del de la pura diversión y que generalmente están ligados a competencias específicas de las materias que se imparten en el contexto para el que fueron diseñados [15, 53]. Además, el desarrollo de un juego para evaluar las competencias concretas de un curso no es una tarea al alcance de cualquier docente. Pero este caso será tratado en el siguiente punto.

#### **Recursos limitados / Coste elevado**

Hay un trabajo que implementa un test psicológico de corrección automática para la evaluación de ciertas competencias genéricas. El diseño de un test de este tipo requiere personal cualificado en psicología, algo que obviamente no tienen todos los docentes. Además, para la experiencia que se encontró en la literatura fue necesario contar con un equipo de expertos que mediante el método Delphi definieron el test, algo con lo que tampoco suelen contar la mayoría de los docentes [5].

En los juegos serios ocurre lo mismo, un docente puede considerar cómo sería un videojuego para que el alumno aplique ciertas competencias genéricas y él después pueda evaluarle en dichas competencias, pero desarrollar un videojuego no es una tarea sencilla al alcance de cualquier docente.

### 3. RESUMEN DE PROBLEMAS ENCONTRADOS

---

#### **Validez de los indicadores**

Por último se han encontrado con un conjunto de trabajos que se basan en el *learning analytics* para evaluar competencias genéricas. Para ello obtienen información de los registros de aprendizaje que utilizan como indicadores asociados a diferentes competencias genéricas [45, 85]. El problema es que los indicadores son fijos, y lo que para un docente puede ser un indicador válido de la competencia de liderazgo, para otro puede no serlo. Por ello, se echa en falta un mecanismo para que sea el propio docente el que seleccione los indicadores que necesita, descarte los que no le sean útiles y los combine hasta obtener el indicador que sea válido para su caso. Es decir, lo que sería deseable es que hubiera un método para que el docente pueda diseñar sus propias evaluaciones en base a estos indicadores.

## 3.2 Requisitos

Antes de comenzar con el desarrollo del método hay que enumerar los requisitos que éste deberá satisfacer para resolver los problemas mencionados en la sección anterior. Los requisitos son los que se detallan a continuación.

### **1. Indicadores objetivos**

Los indicadores utilizados para medir el desempeño de los estudiantes deben ser objetivos. No ha lugar a consideraciones personales o interpretaciones inexactas de rúbricas como ocurre en la autoevaluación o evaluación entre iguales, donde dos evaluaciones de un mismo trabajo realizadas por personas diferentes pueden tener calificaciones diferentes.

### **2. Evaluación escalable**

El método para la evaluación de competencias genéricas deberá ser escalable, de manera que un crecimiento en el número de actividades a evaluar no le suponga un esfuerzo al docente que no pueda abordar. El método deberá estar alineado con las actividades de aprendizaje para que el docente pueda consultar los indicadores con una simple petición a la herramienta.

### 3. Propósito general

El método no debe estar orientado a una competencia específica ni genérica concreta. Es el docente quien diseña sus actividades en el entorno virtual de aprendizaje y el que luego obtiene los indicadores para utilizarlos en la evaluación de la competencia que considere que los estudiantes han desempeñado en dicha tarea (y que queda reflejada en los indicadores).

### 4. Accesibilidad

No deberá ser un requisito que el docente tenga un perfil informático u otro específico para poder realizar las peticiones de los indicadores, ni que contrate a un equipo de expertos para obtener los indicadores. La interfaz en la que se implementará el método debe ser usable y sencilla para que los docentes puedan utilizarla sin requerirles conocimientos de programación, y los formatos a los que se exporte la información serán figuras y documentos en formatos transportables a cualquier hoja de cálculo.

### 5. Diseño de evaluaciones

El método debe poner a disposición del docente los mecanismos necesarios para que éste pueda diseñar sus propios indicadores a partir de los datos objetivos de la actividad de los estudiantes. En el estado del arte nos encontramos con trabajos que obtenían sus evaluaciones a partir de los indicadores del entorno virtual de aprendizaje, pero éstos eran fijos. Es decir, cada competencia se evaluaba con un indicador dado. Pero podía ocurrir que el docente no utilizase las actividades del entorno virtual que proporcionaban dichos indicadores o que plantease las actividades con un enfoque diferente al que realmente tienen. Por ejemplo, uno de los puntos fuertes de un wiki es que favorecen el trabajo colaborativo, y podríamos encontrar herramientas que nos ayuden a valorar el trabajo en equipo de los estudiantes que participan en una página de un wiki mediante indicadores del trabajo colaborativo. Si un docente plantea actividades en el wiki de manera que cada estudiante trabaje individualmente en una página, podrá valorar otras competencias, pero no el trabajo en equipo. Con este método el docente debe ser quien diseñe sus

### **3. RESUMEN DE PROBLEMAS ENCONTRADOS**

---

indicadores, y por tanto, sus evaluaciones a partir de los registros de las actividades de aprendizaje.

# Método para la evaluación de competencias genéricas

En este capítulo se propone un método para la evaluación de competencias genéricas de los estudiantes basado en el diseño de evaluaciones a partir de la actividad de estos en los entornos de aprendizaje. Se describen el método, sus características, la adecuación del método para satisfacer los requisitos deseados y se introduce el tipo de herramienta que se empleará para su implementación.

## 4.1 Introducción

Prácticamente hoy en día los LMS constituyen una pieza fundamental en cualquier contexto en el que se impartan cursos. Mientras que en los cursos virtuales los LMS son el único entorno de trabajo posible, en los cursos presenciales o mixtos, tanto los LMS como otros entornos virtuales de aprendizaje actúan como soporte virtual de las clases, proporcionando multitud de actividades de aprendizaje.

En esta tesis se propone un método de evaluación de competencias genéricas basado en el diseño de evaluaciones (DBA, del inglés, *design-based assessment*) y que tiene su origen en la investigación basada en el diseño (DBR). DBA es un

## 4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

---

método de evaluación de competencias genéricas que se basa en el diseño de indicadores a partir de la actividad generada por los estudiantes en los entornos virtuales de aprendizaje. Los docentes podrán diseñar evaluaciones utilizando estos indicadores y podrán utilizar estas evaluaciones como evidencias del desempeño de competencias genéricas. Si llegado el caso, el docente considera que las evidencias no terminan de reflejar bien el desempeño de las competencias, el método DBA le permitirá refinar los diseños hasta llegar a unas evidencias que sí le sean válidas, o a descartarlas si no llegaran a serlas. Además, los diseños podrán ser utilizados y modificados por otros docentes que busquen adaptarlos a su contexto local o a las competencias genéricas que deseen medir.

En el LMS por ejemplo, los estudiantes acceden a su contenido a diario. Por un lado, habrá estudiantes que accedan al LMS cada día a consultar novedades, participar en foros, subir tareas o descargar apuntes, mientras que por otro lado, habrá estudiantes que entren sólo de manera puntual, a realizar un examen o a subir una tarea. Todas las acciones quedan almacenadas en el registro de actividad de los LMS, y estos registros podrían ser analizados para comprender el proceso de aprendizaje que se está desarrollando mediante técnicas de *learning analytics*. El método DBA se basa en la utilización e integración de técnicas y herramientas de *learning analytics* para la obtención de información del registro de actividad procedente de estos y otros entornos virtuales de aprendizaje con los que diseñar indicadores.

### 4.2 Método: design-based assessment (DBA)

En esta sección se explica el método de evaluación DBA. La sección comienza con un apartado dedicado a la contextualización del método y continúa con la descripción en detalle del mismo.

#### 4.2.1 Contexto

Los estudiantes comienzan a dejar constancia de su actividad en los entornos virtuales de aprendizaje desde el momento en que acceden al entorno. El registro del sistema almacena gran cantidad de información, tanto la participación activa del

## 4.2 Método: design-based assessment (DBA)

---

estudiante, cuando éste interactúa con el entorno virtual de aprendizaje, como la participación pasiva, cuando el estudiante simplemente accede y navega entre sus contenidos.

Los docentes pueden plantear actividades en los entornos virtuales de aprendizaje con la intención no sólo de evaluar ciertas habilidades de los estudiantes, sino también de provocar comportamientos en los estudiantes y ver cómo afrontan ciertas situaciones. Pueden encontrarse patrones de comportamiento en la manera en que los estudiantes abordan ciertas tareas y estos comportamientos podrían ser interpretados como un indicador del desempeño de alguna competencia genérica.

Los guías docentes de las asignaturas incluyen las competencias genéricas de las que los estudiantes deben ser evaluados. Para evaluar una competencia genérica dada, el docente puede diseñar una evaluación a partir de la información relativa a los registros de los entornos de aprendizaje. Aquí comienza un *ciclo de contraste de hipótesis*.

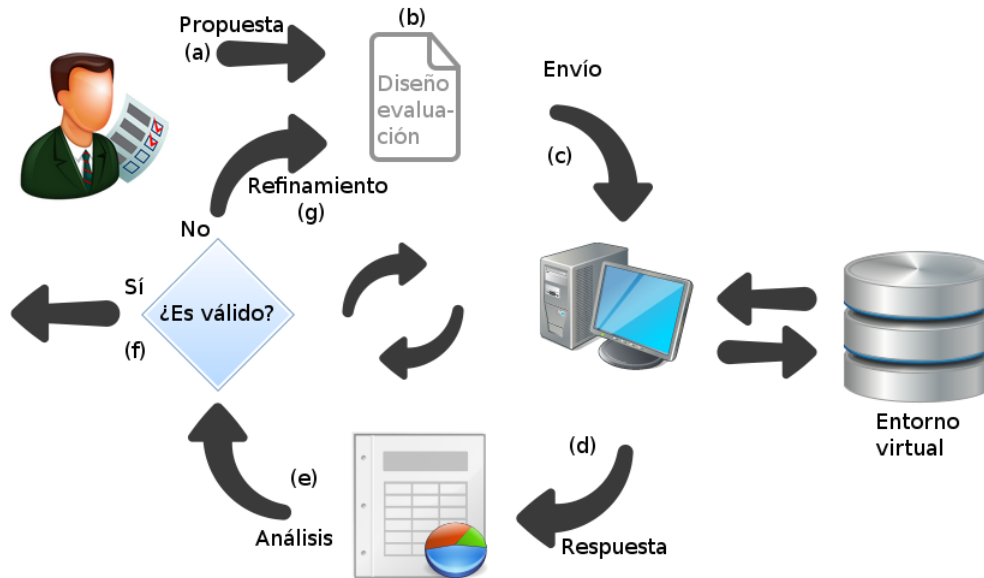
### 4.2.2 Descripción del método

El método DBA es iterativo, ya que permite al docente rediseñar y contrastar hipótesis y resultados hasta confirmar la hipótesis, o por el contrario, descartarla y enunciar una nueva. Podemos decir que el método se integra en un *ciclo de contraste de hipótesis*. Este ciclo consta de una serie de pasos que se muestran en la figura 4.1 y se explican a continuación:

1. *Hipótesis inicial*: el docente formula una hipótesis de partida para la utilización de algún tipo de información de la actividad de los estudiantes en el entorno virtual de aprendizaje para la evaluación de alguna competencia genérica (a). Por ejemplo, se considerará que un estudiante tiene un desempeño correcto de la competencia genérica de planificación y gestión del tiempo si entrega las tareas programadas por el docente en el entorno virtual con anterioridad a la fecha fijada para las mismas.
2. *Diseño y formulación de evaluación*: el docente diseña un indicador para evaluar la competencia a partir de la información del registro y la implementará en la herramienta utilizada para extraer la información (b). Por ejemplo,

## 4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

---



**Figura 4.1:** Diagrama del ciclo de contraste de hipótesis

podríamos considerar partiendo de la hipótesis anterior que un estudiante tendrá un desempeño alto en la competencia de planificación y gestión del tiempo si de las 10 tareas programadas durante el semestre al menos 9 fueron entregadas antes de la fecha fijada para las mismas, un desempeño medio si entregó entre 7 y 8 tareas antes de la fecha fijada y un desempeño bajo si entregó 6 o menos tareas antes de la fecha fijada.

3. *Petición de datos:* se envía la petición de datos al sistema encargado de recuperar la información (c). Las herramientas y su funcionamiento serán explicadas más adelante en la sección 4.3.3.
4. *Validación de resultados:* la herramienta pondrá a disposición del docente los indicadores requeridos (d). El docente los analizará (e) y evaluará si son válidos para el propósito que fueron diseñados, si necesitan ser refinados o si hay que descartarlos (f). En este caso, podrá volver al segundo punto y rediseñar una nueva evaluación (g).

A continuación se presentara un ejemplo para ilustrar el ciclo de contraste de hipótesis.



### Ejemplo de ciclo de contraste de hipótesis

En este ejemplo, el docente pretende obtener indicadores del uso del foro del LMS para evaluar la competencia genérica de las habilidades interpersonales. Durante el curso, el docente planteó una actividad de debate mediante el foro de la asignatura que tuvo lugar durante una semana de las que conforman el curso. El docente aplicará dos iteraciones del ciclo de contraste de hipótesis del método DBA.

### Iteración I

**I.1 Hipótesis inicial.** El docente considera que podría diseñar un indicador válido a partir del número de mensajes que ha escrito cada estudiante en la semana en la que transcurre la actividad y plantea la siguiente hipótesis: *se considerará que un estudiante ha desempeñado satisfactoriamente la competencia genérica de las habilidades interpersonales si ha escrito al menos dos mensajes en el foro en la semana.*

**I.2 Diseño y formulación de la evaluación.** Para diseñar la evaluación, el docente deberá contar con un lenguaje cercano al dominio que le permita formular la consulta que le proporcione los indicadores que requiere la hipótesis inicial. En este ejemplo mostramos la consulta I.1, escrita en el lenguaje SASQL <sup>1</sup> (que se presenta en la sección 5.3.1) y que proporcionará datos sobre la participación de los estudiantes en el foro entre dos fechas.

---

Consulta I.1

```
1 Evidence participacion_foro:
2     get students
3     show participation
4     in forum between 2015-10-21 and 2015-10-27.
```

---

**I.3 Petición de datos.** En este paso, se enviará la consulta a un software que lo interprete, procese y sea capaz de proporcionar los datos solicitados.

---

<sup>1</sup><https://www.asamblea.com/spaces/evalcourse/wiki/SASQL>

## 4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

---

**Tabla 4.1:** Información sobre la participación en el foro de los estudiantes en un periodo concreto de tiempo

id	username	Debate-starter	Debate-participation	Total
1	student1	1	2	3
2	student2	0	4	4
3	student3	0	1	1
4	student4	1	2	3
5	student5	0	2	2

**I.4 Validación de resultados.** El software devuelve al docente los indicadores solicitados mediante el listado 5.2. El docente los analiza y concluye que, en base a la hipótesis inicial, todos los estudiantes menos el 3 (*student3*) habrían desempeñado correctamente la competencia. Sin embargo, el docente considera pobre esta primera aproximación y decide redefinir la hipótesis inicial.

### Iteración II

**II.1 Hipótesis inicial.** El docente plantea la nueva hipótesis inicial de la siguiente manera: *se considerará que un estudiante ha desempeñado satisfactoriamente la competencia genérica de las habilidades interpersonales si ha escrito al menos dos mensajes en el foro y ha interactuado con más de un compañero en la semana.*

**II.2 Diseño y formulación de la evaluación.** El docente diseña la nueva evaluación mediante la consulta SASQL I.2.

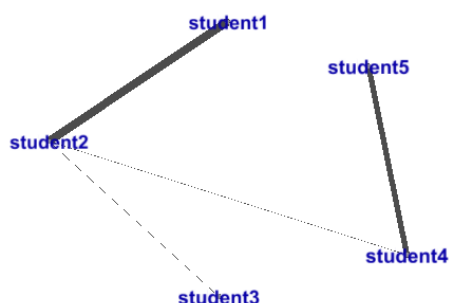
```
_____ Consulta I.2 _____  
1 Evidence interacciones_foro:  
2     get students  
3     show interaction  
4     in forum between 2013-10-21 and 2013-10-27.
```

---

**II.3 Petición de datos.** Se ejecuta la consulta en el software.

## 4.3 Características, requisitos y herramienta

---



**Figura 4.2:** Interacción en el foro en un periodo de tiempo.

**II.4 Validación de resultados.** El software devuelve, entre otros formatos, un grafo que muestra las interacciones (figura 5.7). A tenor de los resultados se observa que sólo dos de los estudiantes cumplen la segunda hipótesis (*student2* y *student4*), y bajo esta nueva hipótesis, ellos serían los dos únicos estudiantes que han mostrado un buen desempeño de la competencia genéricas de las habilidades interpersonales.

### 4.3 Características, requisitos y herramienta

El método DBA para la evaluación de competencias genéricas es un método DBR. Aunque desde el punto de vista del estudiante y considerando la clasificación de métodos de evaluación mostrada en la sección 2.3.2 diríamos que es un método *evaluación formativo*, ya que permite mejorar el aprendizaje mientras este tiene lugar proporcionando información de manera sistemática y continua. Sin embargo, lo consideramos un método DBR ya que no mejora únicamente la acción del alumno, sino que también es un método de investigación que mejora la acción del docente.

Mientras que con respecto a la clasificación de técnicas de evaluación mostrada en la sección 2.3.2, la técnica empleada en este método es la de obtención de *indicadores del trabajo en actividades de aprendizaje*.

#### 4.3.1 Características

En el capítulo de la introducción se indican, partiendo del análisis realizado por Terry Andersen y Julie Shattuck [4], las características que un estudio DBR de

#### 4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

---

calidad debe tener. A continuación se identificará nuestro método DBA en cada una de esas características:

- *Estar situado en un contexto educativo real*: el método DBA se ha creado en un contexto educativo real (la Universidad de Cádiz) para ser utilizado en este y en otros contextos educativos reales. Su validez será tratada en el capítulo de la evaluación.
- *Enfocado en el diseño y prueba de una intervención significativa*: La intervención que se lleva a cabo con el método DBA es un *tipo de evaluación*, siendo este uno de los tipos de intervenciones recogidos en la tipología de intervenciones DBR definida por Andersen [4].
- *Empleo de métodos mixtos*: una intervención DBR implica normalmente la aplicación de diferentes métodos, técnicas y herramientas. El método DBA puede combinarse con otros métodos de evaluación con el fin de contrastar evaluaciones y dar validez a los resultados obtenidos con uno u otro método. Por ejemplo, una aplicación de métodos mixtos podría darse en el ejemplo mostrado en el apartado anterior, donde los estudiantes eran evaluados de sus habilidades interpersonales mediante el método DBA. Esa actividad podría formar parte de un conjunto de actividades de trabajo en equipo configuradas por el docente para realizar una evaluación auténtica. El docente podría combinar los resultados de la aplicación de ambos métodos y así contrastar los resultados.
- *Múltiples iteraciones*: el método DBA permite el diseño de evaluaciones y su contraste de resultados (*ciclo de contraste de hipótesis*). A partir de los resultados, el diseño se puede ir refinando iterativamente hasta que los resultados satisfagan las expectativas del docente o hasta que decida descartarlos por no ser válidos para su evaluación.
- *Asociación colaborativa entre investigadores y docentes*: el proceso de desarrollo de software es cada vez más colaborativo, tratando de integrar todo lo posible a los usuarios finales para avanzar hacia un proceso dirigido por la comunidad donde tanto los actores técnicos como los no técnicos trabajen

### 4.3 Características, requisitos y herramienta

---

juntos para que se cumplan las expectativas. Esto es especialmente apropiado en el campo de los DSL, herramientas informáticas diseñadas para facilitar el desarrollo de software en un contexto específico [61]. Bajo este enfoque se desarrollará el método DBA, colaborando docentes e investigadores en su desarrollo, teniendo así los usuarios finales y expertos en el dominio (los docentes) una participación activa y directa en la definición de una herramienta que implemente el método, es decir, un DSL.

- *Evolución de los principios de diseño*: el método DBA permite la creación de diseños para aplicar a diferentes competencias. Ciertos diseños podrían ser definidos como plantillas o patrones para la evaluación de competencias. Estos patrones podrán ser utilizados por otros docentes, y a partir de su reflexión, modificados para adaptarlos a nuevos contextos. De esta forma, los principios y diseños creados por docentes e investigadores no mueren con una práctica, sino que son puntos de partidas para nuevas teorías y experimentos. Por ejemplo, un docente que trabajase con wikis en sus clases podría utilizar como indicador de desempeño de la competencia de trabajo en equipo el número de contribuciones realizado por cada estudiante a la página de su equipo. Un segundo docente que trabajase con wikis podría mejorar esta evaluación considerando como indicador el peso en bytes de las contribuciones aportadas por cada miembro del equipo. Y un tercer docente podría mejorar esta evaluación añadiendo una restricción temporal a la información aportada al wiki por cada miembro del equipo. Es decir, se parte de un patrón o diseño inicial que los docentes han ido evolucionando para mejorarlo y aplicarlo a su contexto.
- *Comparación con la investigación-acción*: el objetivo del método DBA, al contrario de lo que ocurre con la investigación-acción, no es sólo alcanzar una serie de objetivos a nivel local, sino que también permite evolucionar a nivel teórico y maximizar la generalización, mejorando así la comprensión de aplicaciones prácticas. Además, permite la participación de un equipo formado por investigadores y docentes mientras que, por lo general, la práctica de la investigación-acción es llevada a cabo únicamente por el docente que realiza el experimento. Ambos enfoques son muy parecidos y en ocasiones

## 4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

---

cuesta diferenciarlos. Se podría decir que la investigación-acción carece de ese carácter reflexivo que sí tiene el DBR [27]. Para ilustrarlo utilizaremos el ejemplo de la medición de la competencia de trabajo en equipo mediante el wiki mostrado en el punto anterior. En la investigación-acción, un docente se enfrentaría a esta evaluación por sí solo, sin considerar cómo han evaluado dicha competencia otros en el wiki, sino que adoptaría una solución útil para él en su caso pero que no evolucionaría. Justo al contrario de lo que ocurrió en el ejemplo anterior de wikis, donde cada docente, investigador o equipo de investigadores y/o docentes tenía en cuenta como había sido abordada antes la evaluación del trabajo en equipo en el wiki y cada nueva evaluación se pudo utilizar como base de futuras evaluaciones.

- *Repercusión en las prácticas*: el método DBA puede repercutir directamente en las prácticas. A partir de su aplicación en la docencia práctica y del análisis de resultados, los investigadores y docentes implicados pueden detectar actividades e interacciones que favorecen el desempeño de competencias y que no habían considerado anteriormente. Esto repercutirá directamente en la organización y programación de actividades en los siguientes cursos.

### 4.3.2 Requisitos

En el capítulo 3 se enumeraron los requisitos que el método deberá cumplir. A continuación se indica cómo el método, apoyado en la herramienta que lo implementa, abordan cada uno de estos requisitos.

#### 1. Indicadores objetivos

Los indicadores que se obtienen con el método DBA reflejan valores de los registros de actividad del entorno virtual de aprendizaje. Por tanto, son objetivos en cuanto a que dos estudiantes que tienen los mismos valores en el registro tendrán en el mismo valor en el indicador, y ya será decisión del docente la interpretación de este indicador.

### 2. Evaluación escalable

El método DBA se alinea con las actividades de aprendizaje para que el docente pueda obtener los indicadores del registro para todos o parte de los estudiantes de un curso con una simple petición a la herramienta, sin importar el número de estudiantes.

### 3. Propósito general

El método DBA permite obtener indicadores de la actividad de los estudiantes en el entorno de aprendizaje virtual. Estos indicadores no están asociados a ninguna competencia específica, es información objetiva almacenada directamente en la base de datos del entorno y puesta, mediante la herramienta que implemente el método, a disposición del docente. Por tanto, se puede afirmar que el método DBA es de propósito general pues será el docente quién deba valorar para la evaluación de qué competencia lo podrá utilizar en cada caso.

### 4. Accesibilidad

La información del registro de actividad del entorno virtual está contenida en la base de datos del entorno. Pero ni los docentes suelen tener acceso a dicha base de datos, ni aún teniéndolo tienen que saber utilizar el lenguaje de acceso a bases de datos (SQL) para realizar las consultas a las tablas que contienen la información. Con el fin de poder obtener los indicadores de una forma que sea accesible y utilizable por docentes sin conocimientos de programación, la herramienta que lo implemente debe ser un lenguaje de consultas sencillo que les permita definir las evaluaciones. El lenguaje tiene una sintaxis con términos del dominio de la docencia, de manera que este tipo de docentes pueda utilizarlo. El docente realizará la consulta y, la herramienta ya conectada a la información, le proporciona los valores de los indicadores en diferentes formatos abiertos que el docente podrá visionar y exportar a otras herramientas.

## 4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

---

### 5. Diseño de evaluaciones

El método DBA, mediante el lenguaje de consultas mencionado en el punto anterior, permitirá a los docentes ejercer una labor de diseñador de evaluaciones, utilizando y combinando la información obtenida del entorno de aprendizaje para calcular indicadores que aplicar a las competencias genéricas para las que ellos consideren que les son válidos.

#### 4.3.3 Herramienta

La herramienta que implemente el método debe tener una serie de características que derivan de los requisitos mencionados anteriormente y que son los siguientes:

- *Personalización de evaluaciones*: la herramienta debe proporcionar un mecanismo para permitir al usuario crear sus diseños a partir de los indicadores disponibles en el entorno virtual de aprendizaje. Para permitir al usuario definir los diseños se decide que la herramienta sea un lenguaje de consultas.
- *Cercano al dominio*: el lenguaje debe servir al docente o investigador para definir indicadores a partir de la actividad de los estudiantes en los entornos de aprendizaje, por tanto, debe ser un lenguaje cuya sintaxis esté definida en términos del ámbito educativo.
- *Usabilidad*: el lenguaje debe ser usable por usuarios de perfil no informático, por lo que su utilización debe ser sencilla y entendible por este tipo de usuarios.

A tenor de estos requisitos se decidió que la herramienta informática a implementar sería un *lenguaje específico de dominio* (DSL). Un DSL es un lenguaje de programación orientado a un dominio concreto. Estos se han vuelto muy populares en los últimos años, y una de las principales razones es que facilitan la comunicación con los expertos en el dominio, proporcionando un texto común (código escrito en el DSL) que actúa a la vez como ejecutable y como una descripción que los expertos en el dominio pueden leer para entender como sus ideas se representan en el sistema [48]. En el método DBA, el ejecutable sería el diseño y los expertos en el dominio los docentes.



# Resultados de la investigación

En este capítulo se evalúa si los objetivos de esta tesis se han alcanzado, es decir, la idoneidad del método propuesto para la evaluación de competencias genéricas, así como las herramientas que lo implementan. A lo largo del capítulo se hacen referencias a diferentes competencias genéricas. Para estas referencias se utilizan los identificadores de competencias que se listan en el apéndice D.

El capítulo comienza con la descripción de las herramientas que implementan el método DBA y su ilustración con varios ejemplos. En segundo lugar, se presentan los estudios de caso en los que se ha implementado el método. En tercer lugar, se analizan los cuestionarios de evaluación y sus resultados. En último lugar, el capítulo termina con la discusión y las conclusiones del proceso de evaluación.

## 5.1 Introducción

Para evaluar la idoneidad del método DBA se han implementado tres herramientas que se han utilizado en otros tantos entornos educativos. El trabajo comenzó en el curso 2011-12, cuando se desarrolló una herramienta web para la evaluación de competencias genéricas y específicas a partir del trabajo de los estudiantes en un

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

wiki basado en MediaWiki. Esta aplicación se describe en los antecedentes (sección 5.2). En el siguiente curso se desarrolló una herramienta de tipo DSL para la evaluación de competencias genéricas a partir de la actividad generada por los estudiantes en el LMS (apartado 5.3.1). La última herramienta desarrollada también es de tipo DSL y se utilizó para la evaluación de competencias genéricas a partir de la actividad generada por los estudiantes en un mundo virtual (apartado 5.3.2).

Las herramientas se utilizaron en varias experiencias para la evaluación de competencias genéricas en asignaturas de la Universidad de Cádiz. Estas experiencias se presentan en diferentes casos de estudio que han sido después presentados en congresos y publicados en revistas.

Por último se realizó un cuestionario para que expertos en el dominio evaluaran la idoneidad del método y la herramienta. Ambos fueron presentados a varios colectivos relacionados con la docencia y las TIC, que tras asistir a los cursos y talleres en los que se presentaron, recibieron una invitación para completar el cuestionario de evaluación.

### 5.2 Antecedentes

En los antecedentes se presenta *AssessMediaWiki* (AMW)<sup>1</sup>, una herramienta que se desarrolló para realizar una evaluación cualitativa del trabajo de los estudiantes en el wiki y que surgió para complementar el análisis que realizaba la herramienta *StatMediaWiki* (SMW)<sup>2</sup>. SMW es una herramienta que realiza un análisis meramente cuantitativo de las contribuciones de los estudiantes en el wiki.

El capítulo comienza con un desarrollo de la problemática surgida con los wikis, a continuación se presenta AMW y se finaliza con las conclusiones.

#### 5.2.1 Desarrollo

El uso educacional de los wikis para las experiencias de trabajo colaborativo está en auge debido a las numerosas ventajas que aporta sobre los modelos tradicionales [39]. Algunas de las ventajas sobre los medios tradicionales, ya sean en formato

---

<sup>1</sup><http://assessmediawiki.forja.cica.es>

<sup>2</sup><http://statmediawiki.forja.cica.es>

impreso o en documentos digitales, es que éstos no llevan un registro de ediciones, no facilitan la colaboración distribuida y asíncrona y no pueden ser monitorizados por el docente mientras los estudiantes completan el trabajo.

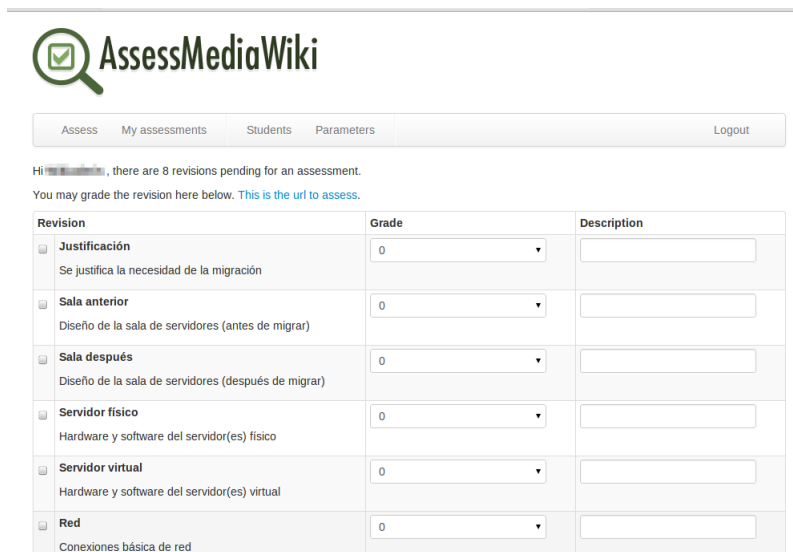
Evaluar el trabajo desarrollado por un grupo de estudiantes en una página del wiki atendiendo a la última versión de la página correspondiente, como se hace con los métodos tradicionales supondría una aproximación muy limitada. Una de las características más interesantes de los wikis es que no sólo almacenan la información de la versión final de cada documento, sino que también almacenan todas las versiones intermedias creadas como resultado de las contribuciones hechas por cada usuario [98]. Esto lo consigue manteniendo un registro con las diferencias entre las ediciones consecutivas de las páginas, registro que se podría utilizar para la obtención de indicadores de diferentes competencias [77]. Las páginas creadas de manera colaborativa podrían ser evaluadas considerando la contribución de cada autor y las dinámicas de grupo en la creación de la página en tiempo real. Por desgracia, realizar una evaluación detallada de cada contribución realizada en el wiki es difícil de abordar por problemas de escalabilidad cuando hay muchos usuarios y éstos participan activamente.

En un proyecto anterior, para poder evaluar el trabajo de los estudiantes en las páginas de un wiki se desarrolló SMW, una herramienta que proporciona al docente información cuantitativa sobre la distribución del trabajo de los estudiantes en las páginas del wiki, es decir, qué parte del trabajo realizado en una página del wiki corresponde a cada estudiante [37]. A partir de esa información cuantitativa se midieron las competencias de la adaptación al cambio, el trabajo en equipo, el aprendizaje y la innovación. Sin embargo, el aspecto cualitativo quedó fuera, ya que el experimento sólo consideró el número, el momento y el tamaño de las contribuciones [79].

Para completar el análisis cuantitativo proporcionado por SMW con un análisis cualitativo y medir el desempeño en otras competencias genéricas se desarrolla AMW. AMW es una herramienta para realizar una evaluación escalable y cualitativa del trabajo realizado en el wiki mediante procedimientos de autoevaluación, evaluación entre iguales y evaluación del docente.

## 5. RESULTADOS DE LA INVESTIGACIÓN

---



The screenshot shows the AssessMediaWiki interface. At the top left is the logo, a magnifying glass over a checkmark, followed by the text 'AssessMediaWiki'. Below the logo is a navigation bar with links for 'Assess', 'My assessments', 'Students', 'Parameters', and 'Logout'. A message states: 'Hi [redacted], there are 8 revisions pending for an assessment. You may grade the revision here below. [This is the url to assess.](#)' Below this is a table with three columns: 'Revision', 'Grade', and 'Description'. The table contains eight rows, each with a checkbox, a title, a description, a grade dropdown set to '0', and a description input field.

Revision	Grade	Description
<input type="checkbox"/> <b>Justificación</b> Se justifica la necesidad de la migración	0	
<input type="checkbox"/> <b>Sala anterior</b> Diseño de la sala de servidores (antes de migrar)	0	
<input type="checkbox"/> <b>Sala después</b> Diseño de la sala de servidores (después de migrar)	0	
<input type="checkbox"/> <b>Servidor físico</b> Hardware y software del servidor(es) físico	0	
<input type="checkbox"/> <b>Servidor virtual</b> Hardware y software del servidor(es) virtual	0	
<input type="checkbox"/> <b>Red</b> Conexiones básica de red	0	

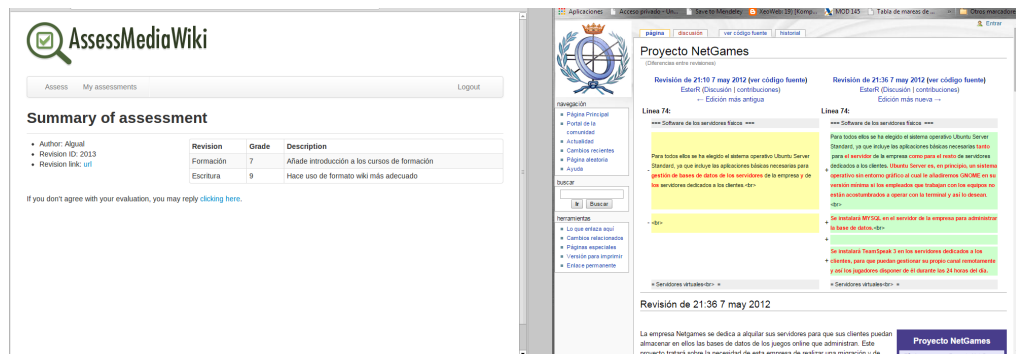
**Figura 5.1:** Una rúbrica de ejemplo en AMW

### 5.2.2 AssessMediaWiki (AMW)

AMW es una aplicación web de código abierto que, al conectarse a una instalación MediaWiki, proporciona procedimientos de autoevaluación, evaluación entre iguales y evaluación del docente, a la vez que mantiene información sobre esas evaluaciones. AMW pone a disposición de los estudiantes una rúbrica previamente definida por el docente para que realicen la evaluación (figura 5.1).

AMW implementa dos roles de usuario distintos: supervisores y estudiantes. Los estudiantes pueden elegir entre distintas opciones: evaluar una revisión, comprobar sus propias aportaciones evaluadas y verificar las evaluaciones ya enviadas. Por otro lado, los supervisores tienen un mayor número de opciones, como definir la rúbrica que los estudiantes deberán completar al realizar sus evaluaciones, modificar los parámetros de los programas o supervisar las evaluaciones que los estudiantes vayan haciendo. AMW implementa una función de selección parcialmente aleatoria. Cuando un estudiante va a realizar una evaluación, el sistema elige automáticamente una de entre el 30% de las más significativas que aún no han sido evaluadas.

Al revisar sus evaluaciones, los estudiantes pueden revisar las notas recibidas y sus justificaciones, así como ver a qué contribución en particular se refiere (figu-



**Figura 5.2:** Ejemplo de retroalimentación formativa y la contribución de wiki evaluada

ra 5.2). Si el estudiante no está de acuerdo con la calificación puede replicar utilizando para ello una rúbrica similar a la que se utilizó en su evaluación, indicando las calificaciones que considera que merece y sus correspondientes justificaciones. Después el docente revisará las propuestas y pondrá la nota definitiva.

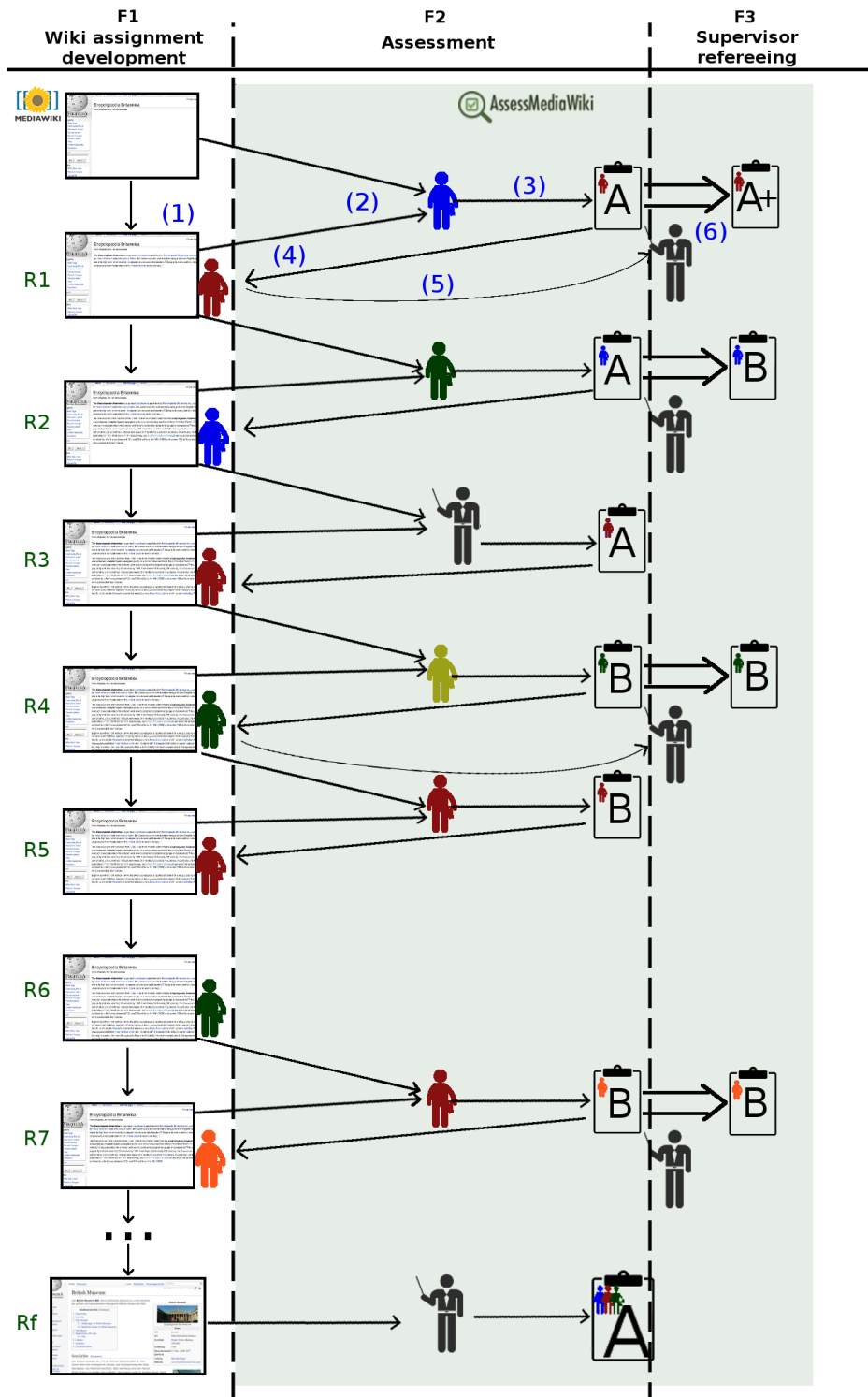
### Ejemplo de uso

El método para la evaluación del trabajo y las competencias desempeñadas en el wiki consta de dos partes: una primera parte en la que lo que se evalúa es el trabajo del wiki (A), basada en procedimientos de autoevaluación, evaluación entre iguales y evaluación del docente; y una segunda parte en la que se evalúan las competencias genéricas (B) y en la que se pone en práctica el *ciclo de contraste de hipótesis*.

#### A. Evaluación del trabajo en el wiki.

El método para la evaluación del trabajo en el wiki se divide también en tres fases: una primera fase en la que los estudiantes realizan sus trabajos en las páginas del wiki (F1), una segunda fase de evaluación (F2) y una tercera fase de revisión del docente (F3). En la figura 5.3 puede verse un diagrama de flujo de trabajo que muestra cada una de las fases del método de evaluación realizado sobre una página del wiki en la que participan varios estudiantes y el docente. A continuación se describen cada una de estas fases.

## 5. RESULTADOS DE LA INVESTIGACIÓN



**Figura 5.3:** Ejemplo de flujo de trabajo para la evaluación cualitativa de un wiki utilizando AMW

**F1: Desarrollo del trabajo en el wiki.** Esta fase se representa en la columna de la izquierda de la figura 5.3, y es en la que los estudiantes realizan el trabajo en las páginas del wiki. Normalmente, cada grupo de estudiantes tendrá que desarrollar su trabajo en una página del wiki. En la zona más alta de la columna se representa el comienzo del trabajo con una página en blanco. El autor de cada contribución se muestra con una figura de color junto a la misma. Para comenzar, el usuario de color rojo crea una página vacía ( $R1$ ). Después, el usuario azul añade contenido a la página ( $R2$ ). En tercer lugar, el usuario rojo modifica de nuevo la página añadiendo texto a la versión dejada anteriormente por el usuario azul ( $R3$ ) y así sucesivamente. Esta fase termina cuando llega la fecha marcada por el docente para que los trabajos estén finalizados ( $R_f$  es la versión final de la página).

Puede verse que, aunque los estudiantes responsables de la página de ejemplo del wiki fuesen el rojo, el azul y el verde, otros estudiantes, como el naranja en la revisión séptima, podrían contribuir a la página del wiki. En ese caso, los miembros del grupo y responsables de la página deben decidir si la contribución debe conservarse, modificarse o eliminarse.

**F2: Evaluación.** Esta fase se muestra en la columna central y comprende las siguientes actividades:

- *Autoevaluación, evaluación entre iguales y evaluación del docente.* Cada contribución es la diferencia entre dos revisiones consecutivas de una página del wiki. Las contribuciones a ser evaluadas se asignan a los estudiantes. El estudiante encargado de evaluar dicha contribución se representa en el gráfico como un usuario coloreado que recibe dos flechas de las revisiones, una del estado de la página antes de la contribución y otra de la revisión que incorpora ya la contribución. Para la evaluación los estudiantes utilizan una rúbrica definida por el docente. Cada contribución sólo se refiere a una contribución concreta de las realizadas a una página del wiki por un único estudiante, por lo que dicha contribución podría ser utilizada como un indicador de la aportación al wiki de dicho estudiante. El estudiante que realizó cada contribución se representa con una figura pegada a la revisión de la página en cada momento. Por ejemplo, en la primera evaluación, se asigna la contribución realizada

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

a la página del wiki por el estudiante rojo (1) al estudiante azul. El estudiante azul comprueba ambas versiones para ver las diferencias entre ambas versiones (2) y realiza la evaluación completando la rúbrica proporcionada por el docente (3). Cabe destacar también otras situaciones interesantes. En la versión R5 de la página se ve que se realiza una autoevaluación, ya que el estudiante rojo, autor de la versión, es el mismo que tiene que evaluar su contribución. Vemos también que en R3 es el docente el que realiza la evaluación de la contribución del estudiante rojo. Esto puede deberse a que el docente manualmente detecta una contribución que considera oportuna evaluar o a que, utilizando la herramienta SMW, detecta un comportamiento extraño en el wiki y quiere contrastar la situación. Puede verse también que hay contribuciones que no reciben evaluación alguna, como ocurre con R6.

- *Revisión de las evaluaciones recibidas.* Los estudiantes pueden revisar las evaluaciones recibidas. Ellos pueden no sólo ver las notas que han recibido con las justificaciones y comentarios que añadieron sus evaluadores, sino también el enlace a la contribución. De esta forma, los estudiantes evaluados reciben una retroalimentación formativa. En la primera de las evaluaciones del diagrama de ejemplo puede verse como el estudiante rojo puede ver su evaluación (4).
- *Réplica.* Si el estudiante evaluado no está de acuerdo con la evaluación recibida tiene la opción de replicarla justificando el motivo de dicha réplica. En el diagrama de ejemplo puede verse como el estudiante rojo considera no apropiada su evaluación y realiza una réplica (5). El docente deberá resolver la réplica en la siguiente etapa.
- *Evaluación final del wiki.* El docente evalúa la versión final de la página del wiki desarrollada por cada grupo de estudiantes. Esta evaluación es recomendable ya que el objetivo principal de la tarea es que los estudiantes realicen un buen trabajo en una página del wiki. Como cualquier otra tarea, deberá ser evaluada por el docente conforme al programa de estudios. Además, algunos de los criterios de evaluación sólo pueden ser evaluados en la versión final de la página, como por ejemplo, la coherencia del texto. De esta forma, aquellas



contribuciones del wiki no evaluadas anteriormente serán ahora implícitamente evaluadas ya que están integradas en el entregable final. Puede verse la evaluación al final del diagrama de ejemplo (*Rf*), y cómo afecta al grupo de estudiantes completo.

El diagrama no recoge algunas situaciones que también podrían darse. Por ejemplo, alguna contribución podría ser evaluada por más de un usuario, ya fuera otro estudiante o el docente. También, para simplificar, el diagrama sólo muestra una calificación para la contribución (A, A+, B, etc.), pero las evaluaciones se realizan con rúbricas multidimensionales.

Un componente interesante de nuestro algoritmo es qué contribución al wiki podrá ser asignada a cada estudiante para su evaluación. Es lo que llamamos *función de selección*, y tiene varios aspectos a tener en cuenta:

- *¿Debería cierta contribución en el wiki ser evaluada por más de un estudiante?* En realidad, tener varias evaluaciones de estudiantes diferentes sobre una misma contribución podría ser interesante para perfeccionar su evaluación y podría proporcionar información al docente para evaluar no sólo al estudiante autor de la contribución, sino también a los evaluadores. De hecho, el número de contribuciones a ser evaluadas es dependiente del objetivo del experimento y su configuración. Cuánto más grande sea el experimento, más contribuciones susceptibles de ser evaluadas tendrá. Sin embargo, el número de evaluaciones que un estudiante puede realizar es limitado (para que siga siendo formativo).
- *¿Qué contribuciones deberían ser evaluadas?* La importancia de evaluar cada contribución puede variar. Por ejemplo, evaluar al menos una mínima cantidad de contribuciones por cada estudiante, página o categoría sería interesante. Pero algunas contribuciones que añadan ciertas características al trabajo pueden ser relevantes o informativas sobre el trabajo realizado por un estudiante. Por ejemplo, aunque las contribuciones que añadan gran cantidad de texto suelen ser más interesantes que las contribuciones pequeñas, una contribución pequeña puede ir relacionada con el cambio de sentido de alguna frase o párrafo. De cualquier forma, un estudiante podría solicitar que

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

una contribución en particular sea evaluada, aunque ésta quede fuera de la función de selección.

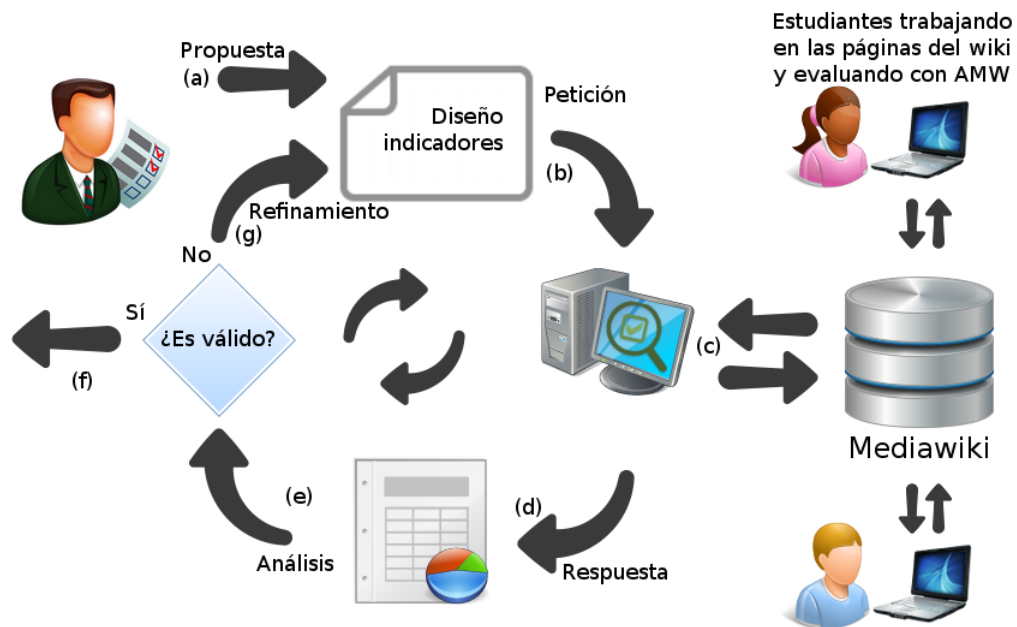
- *¿Quién evalúa cada contribución?* Depende de la importancia que se quiera dar a la autoevaluación, la evaluación del compañero y la del docente. De nuevo, se debería equilibrar el esfuerzo requerido y el detalle a exigir en las evaluaciones.

**F3: Revisión del docente.** En esta última columna se representan dos actividades que corresponden al docente:

- Resolución de las réplicas: el docente revisa las réplicas indicando si proceden o no. En caso de que procedan, modifica la calificación. En el diagrama se puede ver como en la primera contribución, el estudiante rojo realiza una réplica (5) sobre la evaluación reciba por el usuario azul (3). El docente revisa la réplica, la considera apropiada y modifica la calificación (6). En un segundo ejemplo, en la evaluación realizada por el estudiante de color amarillo sobre la contribución realizada por el usuario de color verde puede verse como el docente no acepta la réplica realizada por este último, y mantiene la calificación otorgada inicialmente por el estudiante amarillo.
- Revisión de evaluaciones no replicadas: el docente puede revisar aleatoriamente otras evaluaciones realizadas por los estudiantes que no hayan sido replicadas. En el diagrama puede verse como el docente revisa las evaluaciones realizadas sobre las contribuciones representadas en R2 y en R7, disminuyendo la calificación de la primera y manteniendo la segunda.

### B. Evaluación de competencias genéricas

El método para la evaluación de competencias genéricas se basa en el *ciclo de contraste de hipótesis* (véase sección 4.2). En la figura 5.4 puede verse la descripción del ciclo para el caso de AMW. Esta evaluación se puede llevar a cabo durante o después de que los estudiantes hayan realizado su trabajo en el wiki y sus evaluaciones con AMW. Evidentemente, cuánto más avanzado esté el trabajo más datos habrá para analizar.



**Figura 5.4:** Ciclo de contraste de hipótesis para la evaluación en wikis

El ciclo comienza con el diseño de un indicador por parte del docente que podría plasmar en una hoja de cálculo a partir de los datos que recibirá de las evaluaciones (a). A continuación, el docente realiza la petición a AMW que le proporcionará los datos (b). AMW consulta y procesa los datos en las bases de datos de MediaWiki y AMW (c) y se los envía al docente (d). El docente los integra en la hoja de cálculo previamente diseñada y los analiza (e). Si son válidos para la evaluación finaliza el proceso (f). Por el contrario, si necesitan algún tipo de refinamiento el docente puede rediseñar la evaluación (g) y hacer una nueva petición.

### Indicadores de competencias genéricas

Los indicadores que se mencionan en este punto han sido utilizados en los estudios de caso realizados para este trabajo, pero como se ha mencionado desde un primer momento, este método proporciona indicadores y es el docente el que los utilizará para evaluar las competencias genéricas que considere oportunas.

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

**Trabajo en equipo (TMW)** El indicador considerado para el TMW es el *ratio de miembros del equipo que trabajaron en un mismo criterio*. La rúbrica que utilizan los estudiantes para evaluar se compone de un conjunto de criterios. Cada criterio puede hacer referencia a una parte del trabajo. En todas las ediciones de un wiki no se trabajan en las mismas partes del trabajo, por lo que al ser evaluado, un estudiante puede tener nota en unos criterios y no tenerla en otros. Si más de un estudiante ha trabajado en la misma parte de una página wiki y su aportación ha sido significativa, tendrán nota en dicho criterio. Por tanto, partiendo de la cantidad de criterios que tiene un trabajo y del ratio de miembros del equipo que ha trabajado en cada criterio tendremos un indicador del TMW.

**Comunicación y aplicación del conocimiento (COM y KNO)** El indicador considerado para la COM y KNO es la *media de las notas recibidas por todos los miembros del grupo*. Este indicador mide la incidencia que tuvieron las contribuciones realizadas en el éxito del proyecto. Una calificación pobre en una contribución puede significar que alguna contribución wiki obtuvo una buena nota en un cierto criterio de la rúbrica pero una mala nota en el otro (el autor de la contribución soluciona un problema y crea uno nuevo). Probablemente, esto se debió a una mala comunicación entre los miembros del equipo o poco compromiso de un determinado estudiante en el objetivo global del grupo.

**Calidad del trabajo producido (QUA)** El indicador considerado para la QUA es la *media de las notas que cada estudiante individualmente recibió*. Unas calificaciones altas en las evaluaciones recibidas pueden significar que el trabajo que el estudiante está produciendo es de calidad. Si el estudiante produjese mucho contenido, pero este no fuese de calidad, las calificaciones no serían buenas. Es decir, sus calificaciones están teniendo en cuenta el aspecto cualitativo del trabajo y por tanto una nota alta significaría un trabajo de calidad.

**Capacidad crítica (CRI)** El indicador considerado para la CRI es el *número de evaluaciones que el estudiante realizó con respecto al número de dichas evaluaciones cuya nota fue modificada por el docente*. Este indicador mide la competencia

## 5.2 Antecedentes

COMP.	INDICADORES CUALITATIVOS (AMW)	INDICADORES CUANTITATIVOS (SMW)
TMW	Ratio de miembros del equipo que trabajaron en un mismo criterio (en la misma o distinta página del grupo)	Ratio de miembros del equipo que contribuyeron a una misma página del wiki en las páginas de su proyecto
COM - KNO	Media de las notas recibidas por todos los miembros del grupo	Porcentaje de miembros del equipo que contribuyeron al menos a un 20% del trabajo realizado
QUA	Media de las notas que cada estudiante individualmente recibió	Tamaño de la contribución individual en bytes
CRI	Número de evaluaciones que el estudiante realizó con respecto al número de dichas evaluaciones cuya nota fue modificada por el docente	No considerada

**Tabla 5.1:** Resumen de las competencias evaluadas para cada tipo de indicador

de un estudiante para evaluar el trabajo hecho por otros. Si recibiera un número fijado de réplicas en sus revisiones y estas fueran revisadas por el docente modificando las calificaciones, podríamos considerar que dicho estudiante no ha desempeñado bien dicha competencia.

En la tabla 5.1 puede verse una comparación entre los indicadores considerados a partir de la evaluación cualitativa que se realizaría con AMW y los que se obtienen a partir de la evaluación cuantitativa realizada con SMW.

### Conclusiones

AMW es una herramienta de evaluación asistida que proporciona un análisis cualitativo interesante, pues al análisis cuantitativo proporcionado por SMW, se podría añadir una información cualitativa que ayudase a detectar hábitos no deseados de los estudiantes. Por ejemplo, la contribución de un estudiante que copiase y pegase

## **5. RESULTADOS DE LA INVESTIGACIÓN**

---

una gran cantidad de texto en una página del wiki sería considerada a efectos cuantitativos como una contribución de peso. Pero desde un punto de vista cualitativo, podría detectarse que la contribución había sido plagiada de Internet.

Sin embargo, en la aplicación de AMW nos encontramos con que se dan algunos de los problemas detectados en la revisión de la literatura para este tipo de herramientas de evaluación asistida. En primer lugar, delegar la evaluación en los estudiantes genera situaciones de subjetividad, pues pueden existir apartados en la rúbrica que los estudiantes interpreten o valoren de forma diferente. Y en segundo lugar, como vemos en el diagrama, el profesorado también revisa las evaluaciones de sus estudiantes, y más aún si estos replican, con lo que los problemas de escalabilidad comentados en la sección 3.1.1 también podría darse.

### **5.3 Lenguajes de dominio**

Después de analizar la experiencia llevada a cabo con la herramienta AMW, aparecieron algunos de los problemas encontrados en la revisión de la literatura, por lo que se decidió considerar otro tipo de herramienta para la implementación del método. Esta herramienta debía no sólo evitar dichos problemas, sino cumplir los requisitos de cercanía al dominio, personalización de las evaluaciones y de usabilidad (subsección 4.3). A tenor de estas condiciones se optó por la implementación de un lenguaje de programación orientado al dominio (DSL), que tuviera una sintaxis sencilla y que diese a los usuarios la posibilidad de diseñar sus propias evaluaciones.

A continuación se presentan los dos DSL que se han desarrollado para la implementación del método, cada uno aplicable a un contexto diferente. En el apartado 5.3.1 se presenta el primer DSL, orientado al cálculo de indicadores de los entornos virtuales de aprendizaje, mientras que en el apartado 5.3.2 se presenta el segundo DSL, orientado a los mundos virtuales.

#### **5.3.1 EvalCourse y SASQL**

El LMS es el núcleo de los cursos virtuales, donde los estudiantes acceden para interactuar con el curso. La interacción de los estudiantes genera gran cantidad de

información que queda registrada en el LMS. Esta interacción será la base para los indicadores del desempeño de competencias genéricas.

### Descripción

Para diseñar evaluaciones a partir de los indicadores del LMS se define *Simple Assessment Specific Query Language* (SASQL). SASQL es un lenguaje formal para la ejecución automática de consultas simples escritas utilizando un lenguaje específico de evaluación. SASQL tiene una sintaxis simple, con términos del dominio docente y orientado a la evaluación de competencias genéricas [13]. De esta forma, los docentes pueden fácilmente diseñar indicadores a partir de la actividad registrada en el LMS sin requerir conocimientos técnicos en bases de datos o programación informática.

También se implementa *EvalCourse*, una herramienta informática que ejecuta instrucciones escritas en SASQL, proporcionando como resultado los valores para los indicadores solicitados. EvalCourse se comunica con el LMS para extraer la información del registro de actividad. EvalCourse<sup>1</sup> está basado en el IDE de la plataforma Eclipse, fue implementado utilizando Xtext [42] dentro del Eclipse Modeling Framework y está disponible como software libre bajo licencia GNU GPL.

El desarrollo de EvalCourse y SASQL se basa en los principios y técnicas de la ingeniería dirigida por modelos (MDE, del inglés Model-Driven Engineering). Este enfoque promueve la construcción de artefactos software de un modo flexible y rápido mediante el desarrollo de modelos y sus transformaciones. Nuestro DSL se define en términos de su sintaxis abstracta o metamodelo, su sintaxis concreta y un conjunto de plantillas para la transformación de modelos de consulta en código ejecutable dependiente del LMS.

El metamodelo de SASQL puede verse en la figura 5.5. La entidad principal es el indicador o evidencia (*evidence*). Esta evidencia se aplicará a una herramienta (*tool*) que puede ser tarea (*assignment*), foro (*forum*), campus (*campus*), taller (*workshop*) o wiki (*wiki*), en las que se observará un indicio (*sign*), que puede ser participación (*participation*), entregas (*milestones*), accesos (*access*), evaluaciones

---

<sup>1</sup><https://www.asamblea.com/spaces/evalcourse>

## 5. RESULTADOS DE LA INVESTIGACIÓN

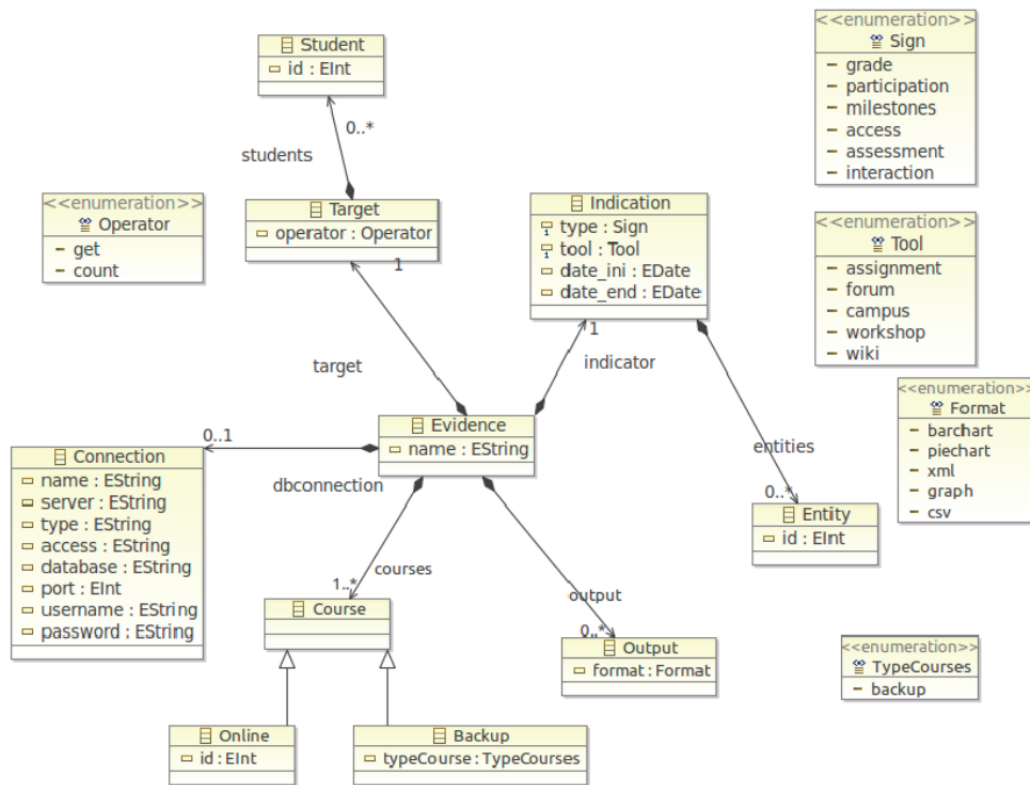


Figura 5.5: Metamodelo de SASQL



(*assessment*), interacción (*interaction*) o calificación (*grade*). Además, puede actuarse sobre una actividad específica (*entity*) o sobre todas las actividades de un tipo que se han dado en el LMS. La conexión a la base de datos del curso se declara en la entidad *connection*.

En el siguiente código puede verse la sintaxis de SASQL. En la primera línea se comienza con la palabra reservada *Evidence* seguida del nombre que se dará al indicador. Ese nombre será el que tengan todos los ficheros de salida. En la segunda línea se escriben los términos obligatorios *get students*. En la tercera, se indica qué indicio se quiere extraer (*show milestones — participation — access — interaction — assessment — grade*). En la consulta, la tercera línea se divide en dos (tercera y cuarta) para que puedan ser visualizadas todas las opciones. En la quinta, se indica sobre qué herramienta se quiere obtener el indicio (*in assignment — forum — campus — workshop — wiki [list of ids]*). También la quinta línea se divide en dos (quinta y sexta). En la séptima línea, que es opcional, se indica el rango de fechas sobre los que se extraerá la información. Y por último, en la octava se especifica si la conexión se realizará directamente a la base de datos o sobre una copia de seguridad almacenada en un fichero.

---

Sintaxis de SASQL

```
1 Evidence indicator_name:
2 get students
3 show milestones | participation | access
4           | interaction | assessment | grade
5           | wiki [list of ids]
6 between YYYY-MM-DD and YYYY-MM-DD
7 from course id | backup.
```

---

El objetivo es que EvalCourse pueda ser utilizado para cualquier LMS, pero las primeras versiones han sido implementadas para funcionar en Moodle<sup>1</sup>. Moodle es un sistema de gestión de aprendizaje de código abierto con más de 64.000 sitios registrados<sup>2</sup>.

---

<sup>1</sup><https://moodle.org>

<sup>2</sup><https://moodle.net/stats>

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

### Ejemplo

Para evaluar las competencias genéricas el docente deberá definir los indicadores que serán extraídos de la actividad de cada estudiante en el LMS. Se ilustra el método a partir de un ejemplo de uso de EvalCourse y su ejecución con los foros del LMS. Los foros suelen ser una de las herramientas incorporadas por los LMS para la interacción entre los estudiantes. Es evidente que la comunicación oral es una manera muy rica de comunicarse, que proporciona múltiples signos no verbales como las expresiones faciales o el tono de voz. En contraste, las comunicaciones escritas proporcionan otras ventajas. Una de las más importantes para la educación es que el estudiante dispone de un tiempo para la reflexión. Por esta razón, podría preferirse la comunicación escrita a la oral cuando se busca un aprendizaje cognitivo [49].

La obtención de indicadores se basará en el *ciclo de contraste de hipótesis* (figura 5.6). En primer lugar, es necesario que los estudiantes hayan interactuado en el LMS, de manera que su actividad haya quedado registrada. Entonces, el docente propone un diseño de evaluación mediante una consulta SASQL (a) y envía esta consulta a EvalCourse (b). EvalCourse procesa la consulta, realiza la petición a la base de datos y recoge los datos (c). Entonces EvalCourse devuelve los resultados (d). El docente analiza los resultados conforme a su propuesta de evaluación de competencias (e), terminando el proceso si éstos son válidos para él (f). Por el contrario, si los resultados no son válidos como indicadores de la competencia, entonces el docente podrá rediseñar la evaluación (g). En cualquier caso, el docente podrá reutilizar el diseño cuántas veces sea necesaria a lo largo del curso y monitorizar la evolución de los indicadores de cada estudiante.

En este ejemplo, el docente pretende obtener indicadores del uso del foro para evaluar la competencia genérica de las habilidades interpersonales (INT). Durante el curso, los estudiantes interactuarán en el foro conforme a las instrucciones proporcionadas por el docente. Cuando lo desee, el docente podrá utilizar EvalCourse para obtener los indicadores.

El docente considera que podría diseñar un indicador válido a partir del número de mensajes que ha escrito cada estudiante en un período de tiempo particular y plantea la siguiente hipótesis: *se considerará que un estudiante ha desempeñado satisfactoriamente la competencia genérica de las INT si ha escrito al menos dos*

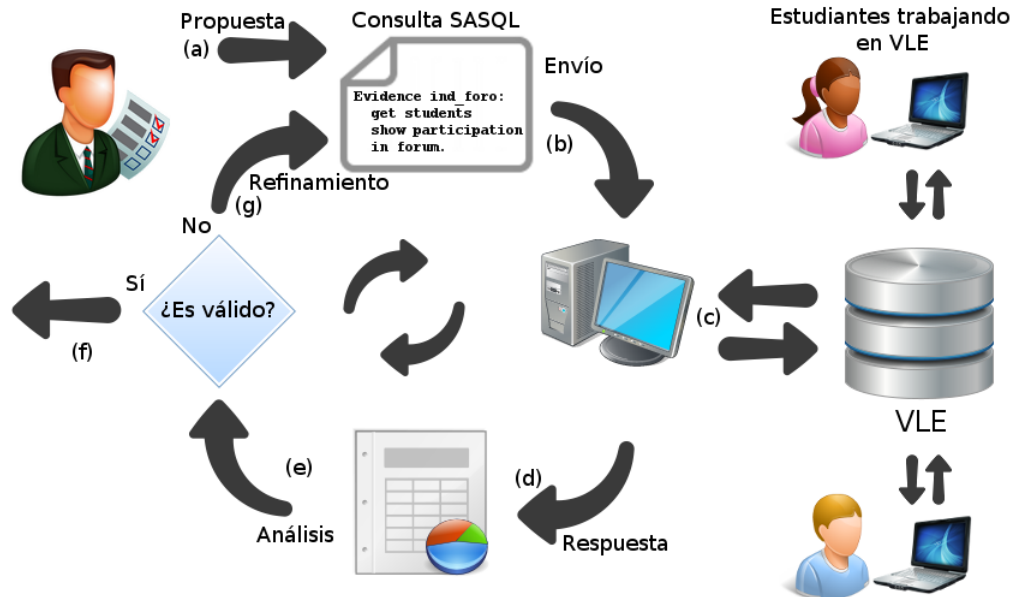


Figura 5.6: Ciclo de contraste de hipótesis utilizando EvalCourse

*mensajes en el foro en la semana.* Para ello se diseña la consulta que se muestra a continuación (consulta SASQL participaciones foro). EvalCourse procesa la consulta y devuelve los resultados que se pueden ver en la tabla 5.2.

```

1 Evidence participacion_foro:
2     get students
3     show participation
4     in forum between 2015-10-21 and 2015-10-27.
    
```

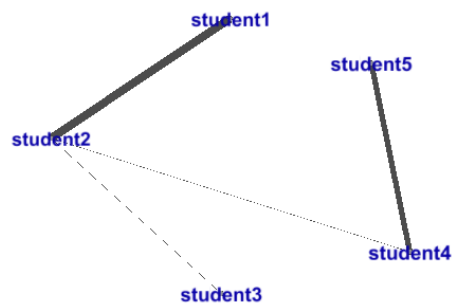
A la vista de los resultados, y en base a la hipótesis inicial, se podría decir que todos los estudiantes menos el 3 (*student3*) habrían desempeñado correctamente la competencia. Sin embargo, el docente considera que esta primera aproximación es un poco pobre y decide completar su hipótesis de la siguiente manera: *se considerará que un estudiante ha desempeñado satisfactoriamente la competencia genérica de las INT si ha escrito al menos dos mensajes en el foro y ha interactuado con más de un compañero en la semana.* Para ello escribe la consulta SASQL interacciones foro.

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

**Tabla 5.2:** Información sobre la participación en el foro de los estudiantes en un período concreto de tiempo

id	username	Debate-starter	Debate-participation	Total
1	student1	1	2	3
2	student2	0	4	4
3	student3	0	1	1
4	student4	1	2	3
5	student5	0	2	2



**Figura 5.7:** Interacción en el foro en un período de tiempo.

```
_____ Consulta SASQL interacciones foro _____  
1 Evidence interacciones_foro:  
2     get students  
3     show interaction  
4     in forum between 2013-10-21 and 2013-10-27.
```

Además del listado con las interacciones, EvalCourse proporciona varias figuras con la representación de la información. Para esta última consulta se devuelve un grafo para una mejor visualización de las interacciones (figura 5.7). A tenor de los resultados vemos que sólo dos de los estudiantes cumplen la segunda hipótesis (*student2* y *student4*).

De esta manera, el docente irá redefiniendo sus hipótesis hasta dar con los indicadores que a su juicio satisfagan la evaluación de la competencia genérica.

### Propuesta de indicadores para el ejemplo

En este apartado se muestra una propuesta de uso de indicadores obtenidos mediante EvalCourse para la evaluación de competencias genéricas. Al igual que ocurre con el resto de herramientas, estos indicadores podrían ser válidos para unos docentes y no serlos para otros, así como que habrá otras muchas formas de combinar la información del registro para obtener indicadores válidos para otras competencias genéricas.

### Habilidades interpersonales (INT)

Para la evaluación de esta competencia genérica se propone utilizar la participación en el foro. Al crear grupos de trabajo en el LMS se puede crear un foro para cada grupo. Durante el curso se fomenta que los estudiantes intervengan en dichos foros para comunicarse entre los miembros del equipo y dejar constancia de los mensajes de cara al docente. Por tanto, si no utilizan el foro, los estudiantes no tienen calificación en esta competencia. Por ejemplo, podría fijarse que un estudiante con tres o más intervenciones en el foro tuviera una evaluación positiva en la competencia. Para obtener la información que nos permita utilizar este indicador utilizaríamos la consulta mostrada a continuación, que devolvería como resultado el listado con los mensajes iniciados, las respuestas dadas y el total de mensajes de cada estudiante en el foro, siendo este último valor el que utilizaríamos como indicador positivo para aquellos estudiantes que tengan tres o más participaciones.

---

```
1 Evidence habilidades_interpersonales:  
2     get students  
3     show participation  
4     in forum.
```

---

### Liderazgo (LEA)

Para evaluar la competencia LEA de los estudiantes se proponen también los foros creados para la comunicación de los equipos de trabajo. Para ello se podría considerar la cantidad de debates que cada estudiante ha iniciado. Por ejemplo, un

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

estudiante que iniciara dos o más debates tendría una evaluación positiva en la competencia LEA. Para obtener la información que nos permita utilizar este indicador utilizaríamos la consulta mostrada a continuación, que devolvería como resultado el listado con los mensajes iniciados, las respuestas dadas y el total de mensajes de cada estudiante en el foro, siendo el primer valor el que utilizaríamos como indicador positivo para aquellos estudiantes que tengan dos o más participaciones (véase como ejemplo el listado mostrado anteriormente en la tabla 5.2).

---

```
1 Evidence liderazgo_foro:  
2     get students  
3     show participation  
4     in forum.
```

---

El wiki de Moodle también se podría prestar a la evaluación de la competencia genérica de LEA. Si tenemos registros de un usuario que se encarga de las primeras ediciones de una página del wiki (organizando el trabajo del equipo) y registros de actividad de ese usuario tras los registros de actividad de sus compañeros (revisando el trabajo de estos), podría considerarse que ese usuario está ejerciendo una labor de líder en el equipo. Para obtener la información que nos permita utilizar este indicador utilizaríamos la consulta mostrada a continuación, que devolvería como resultado el listado con el momento y el peso de las contribuciones de cada estudiante. El docente utilizaría estos valores contrastaría los valores de los otros estudiantes para valorar si ha ejercido o no de líder.

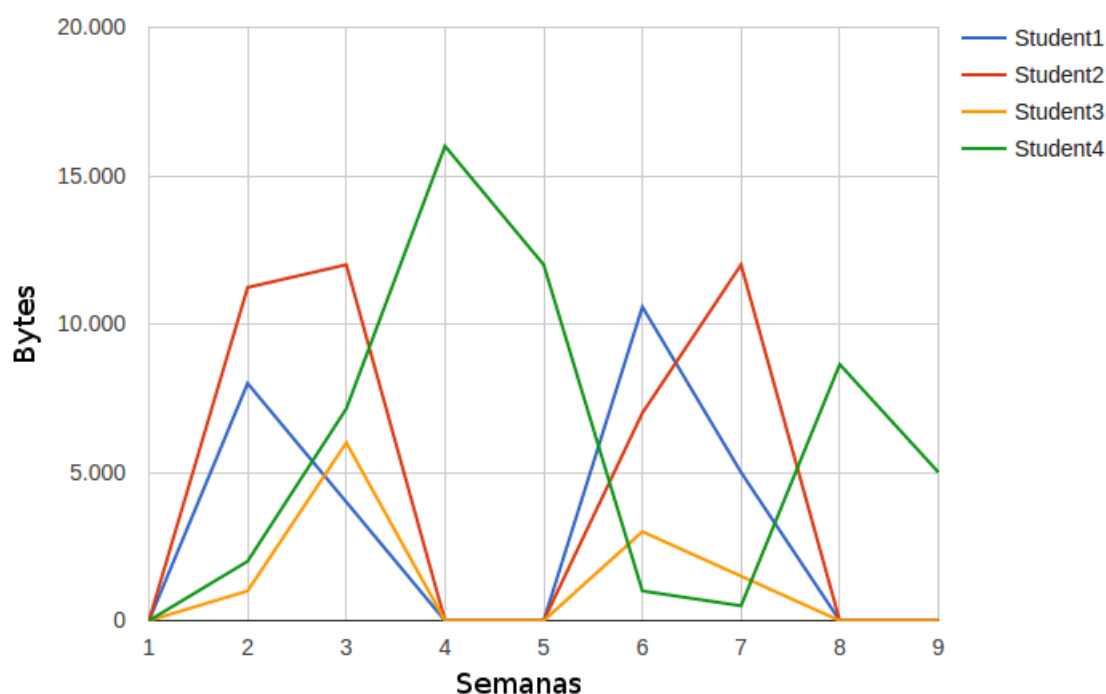
---

```
1 Evidence liderazgo_wiki:  
2     get students  
3     show interaction  
4     in wiki.
```

---

Además del listado, la consulta devuelve un gráfico con las contribuciones realizadas semanalmente por cada estudiante a la página del wiki. A partir del ejemplo mostrado en la figura 5.8, se muestra cómo podría interpretarse dicha figura para

### 5.3 Lenguajes de dominio



**Figura 5.8:** Gráfico de contribuciones semanales de cada estudiante a la página del wiki utilizable para evaluar la competencia LEA

evaluar en este caso la competencia LEA. Puede verse como durante 9 semanas los estudiantes han estado trabajando en su página del wiki. Dicho trabajo tenía 2 entregas, una en la semana 5 y otra en la semana 9. Puede verse como antes de dichas semanas todos los estudiantes trabajan en la página. Sin embargo, puede verse también como el estudiante 4 (*Student4*) es el único que realiza contribuciones justo antes de la entrega. Esto podría ser interpretado por el docente como una labor de coordinación y de ajuste del trabajo realizado por el equipo en los días previos. Este estudiante probablemente se ha dedicado a dar coherencia al texto creado por el resto del equipo, así como a corregir errores para dejar el trabajo apunto para la revisión del docente.

#### **Pensamiento crítico (CRI)**

En Moodle hay una actividad que son los talleres (*workshops*). En esta actividad, los estudiantes tienen que entregar un ejercicio conforme a las instrucciones del

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

docente que podrá ser autoevaluada o evaluada por uno o varios compañeros. Por ejemplo, podría plantearse que cada estudiante tuviera que hacer varias tareas. Una vez entregada cada tarea, el docente pondría la solución del ejercicio a disposición de los estudiantes, y cada tarea sería evaluada por dos compañeros y por el propio estudiante. Para evaluar la competencia CRI, el docente utilizaría para cada tarea la diferencia entre la media de las notas dadas por sus compañeros y la que se asignó el propio estudiante. Para obtener la información que permita utilizar este indicador se usará la consulta mostrada a continuación, que devolvería como resultado el listado de estudiantes con tres columnas adicionales, la primera con la calificación fruto de la autoevaluación y las dos siguientes de la evaluación realizada por los otros dos estudiantes (véase el ejemplo mostrado en la tabla 5.3). El docente utilizaría la diferencia entre estos valores para valorar la competencia genérica.

**Tabla 5.3:** Listado de estudiantes generado por la consulta para la evaluación del CRI

id	username	auto	peer1	peer2
1	student1	8	9	8
2	student2	7	4	5
3	student3	9	9	8
4	student4	10	7	9
5	student5	10	8	10
6	student6	9	9	8
7	student7	8	9	8
8	student8	6	5	7

---

```
1 Evidence pensamiento_critico:
2     get students
3     show assessment
4     in workshop.
```

---



### Planificación y gestión del tiempo (PLA)

Las tareas programadas en el LMS tienen una fecha límite de entrega. Sin embargo, el docente puede configurar la actividad para que permita envíos retrasados. El docente podría establecer un número mínimo de trabajos entregados antes de la fecha límite para considerar que gestionó bien el tiempo. De esta manera, un docente podría considerar que si las tareas se han entregado tarde, pero son correctas, tengan una buena calificación en lo que a las habilidades y conocimientos específicos que requería la tarea. Pero con respecto al desempeño de la competencia PLA llevada a cabo por ese estudiante la calificación sería negativa, dado que ha incumplido la planificación y los plazos que se acordaron a comienzos del curso. Para obtener la información que nos permita utilizar este indicador se usará la consulta mostrada a continuación, que devolvería como resultado el listado de estudiantes indicándose para cada uno la cantidad de tareas entregadas, las que ha entregado a tiempo de estas y las que han quedado pendientes. A partir de esta información el docente podrá evaluar la competencia PLA de sus estudiantes.

---

```
1 Evidence planificacion:  
2     get students  
3     show milestones  
4     in assignment.
```

---

### Resumen

En la tabla 5.4 se muestra el resumen con los indicadores propuestos para la evaluación de competencias genéricas a partir del uso de la herramienta EvalCourse y el lenguaje SASQL. Algunos de estos indicadores fueron utilizados en la experiencia presentada en [10].

#### 5.3.2 EvalSim y VWQL

La segunda propuesta de DSL se aplica a los mundos virtuales. Aunque muchos investigadores han reconocido la potencia educativa y motivacional de los videojuegos, hay pocos estudios empíricos que hayan investigado recientemente su impacto

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

COMPETENCIA	ACTIVIDAD DEL LMS	INDICADOR
INT	Foro	Número mínimo de intervenciones
LEA	Foro	Conversaciones iniciadas en los foros del equipo de trabajo
LEA	Wiki	Contribuciones a la página justo antes de la entrega final
CRI	Taller	Diferencia entre la media de las calificaciones recibidas por el estudiante y la que él mismo se asignó
PLA	Tarea	Entrega antes de la fecha límite

**Tabla 5.4:** Resumen de propuesta de indicadores a partir del uso de EvalCourse

en el aprendizaje de los estudiantes [18]. Lo mismo se puede decir de los entornos de aprendizaje como los mundos virtuales (Second Life, OpenCobalt, etc.) [58]. Esto se debe a que normalmente no se distribuyen bajo licencia libre ni ofrecen API de consulta y, por consiguiente, los docentes difícilmente pueden analizar las interacciones de los estudiantes o analizar su impacto en el aprendizaje y en los resultados de aprendizaje [31, 73].

Ante esta situación, en un trabajo previo desarrollamos nuestro propio mundo virtual, basado en OpenSim (software de código abierto) [19]. De esta forma se podría evaluar la competencia de la comunicación en lengua extranjera (LAN) a partir de las interacciones llevadas a cabo por los estudiantes en el mundo virtual. A continuación se explicará cómo.

### Descripción

Para este trabajo se creó VWQL (*Virtual Worl Query language*). VWQL es un DSL creado para diseñar indicadores a partir de la actividad registrada en un mundo virtual basado en OpenSim. EvalSim es el sistema que procesa las consultas de VWQL. Ha sido desarrollado bajo un enfoque MDE para modelar procesos para el cálculo de los indicadores que se requieran.

## 5.3 Lenguajes de dominio

---

La sintaxis del lenguaje (versión beta 0.1) puede verse en el cuadro de sintaxis VWQL. La primera línea especifica el nombre del indicador (*name\_of\_the\_indicator*) y se utiliza para diferenciar los diferentes archivos producidos por EvalSim. La segunda línea comienza obligatoriamente con *get students* y a continuación se ha de especificar si se quiere obtener la información para todos los estudiantes que participaron en la experiencia o sólo para algunos de los que participaron (indicando sus identificadores numéricos de usuario). La última línea indica el tipo de información a extraer tras el término obligatorio *show*, y esta información puede ser de alguno de los siguientes tipos:

- *words*: número de palabras escritas en el chat de texto. Por defecto cuenta todas las palabras, a no ser que se indique un diccionario, en cuyo caso únicamente muestra, de las palabras introducidas en el chat de texto, las palabras en dicho diccionario.
- *sentences*: número de frases escritas en el chat de texto.
- *single*: número de intervenciones de una sola palabra escritas en el chat de texto.
- *turns*: número de turnos empleados en chat de texto. Un turno es un conjunto de frases consecutivas escritas por el mismo usuario.
- *time*: número de minutos jugados en el mundo virtual.
- *points*: número de puntos obtenidos en el mundo virtual. La manera en que los puntos se obtengan dependerá específicamente del mundo virtual jugado.

---

Sintaxis VWQL

```
1 Evidence name_of_the_evidence:
2   get students [id_of_the_student]
3   show ( words [dict] | sentences | turns | single |
4         | time | points )+
5   between YYYY-MM-DD and YYYY-MM-DD
```

---

## 5. RESULTADOS DE LA INVESTIGACIÓN

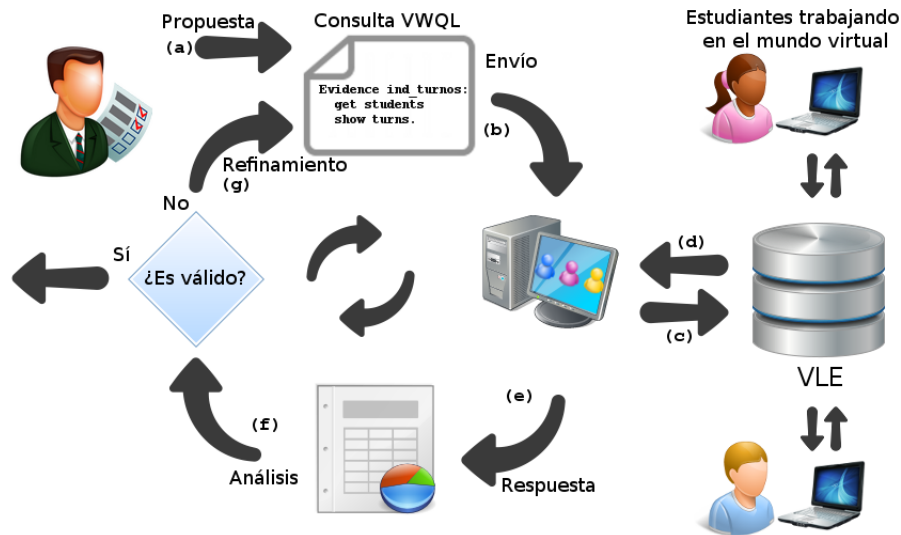


Figura 5.9: Ciclo de contraste de hipótesis con EvalSim

### Ejemplo

En el juego los estudiantes interactúan entre ellos por medio de un chat de texto. Las conversaciones quedan almacenadas en la base de datos de EvalSim. Utilizando el *ciclo de contraste de hipótesis* los docentes definirán indicadores a partir de la actividad sus estudiantes (figura 5.9).

El docente propone una evaluación (a) que bajo su criterio le vaya a proporcionar indicadores válidos para evaluar alguna competencia genérica y la envía a EvalSim (b). EvalSim procesa la petición y realiza la solicitud a la base de datos del mundo virtual (c). Una vez que recibe los datos (d), los transforma y devuelve los resultados debidamente formateados al docente (e). El docente los analiza, y si considera que son indicadores válidos para evaluar la competencia (f) termina el ciclo. Sin embargo, si entiende que no les sirven o considera que debería refinarlo podría volver a diseñar una nueva evaluación (g), reiniciándose de nuevo el ciclo.

En este ejemplo se parte de que el mundo virtual se está utilizando en una asignatura de idioma alemán. En un ejercicio práctico confeccionado por el docente, los estudiantes han de participar en un juego de roles (*role-play*) y son advertidos de que deben expresarse adecuadamente en sus respuestas a fin de aplicar los recursos lingüísticos aprendidos en clase.

**Tabla 5.5:** Información sobre las respuestas de una sola palabra dadas por los estudiantes en el role-play

id	username	singles
1	student1	0
2	student2	1
3	student3	4
4	student4	0
5	student5	2
6	student6	4
7	student7	3
8	student8	1

Una vez que todos los estudiantes han participado el juego, el docente decide utilizar como indicador el número de intervenciones de una sola palabra utilizada por los estudiantes. De esta manera, todos aquellos estudiantes que hayan utilizado una sola palabra para responder a alguna pregunta tendrán una evaluación negativa en el desempeño de la competencia correspondiente. Puede verse el código empleado en la consulta VWQL 1.

```
_____ Consulta VWQL 1 _____  
1 Evidence respuestas_cortas:  
2   get students  
3   show single.
```

---

Como resultado a la consulta, el sistema devuelve el listado que puede verse en la tabla 5.5. A la vista del mismo, el docente considera que hay situaciones en las que una respuesta corta podría admitirse como válida, por lo que decide aceptar hasta dos respuestas cortas en la participación de cada estudiante. Ante este enfoque, el docente considera que los estudiantes 3, 6 y 7 (*student3*, *student6* y *student7*) tuvieron un desempeño bajo en la competencia al haber empleado cuatro, cuatro y tres respuestas cortas respectivamente.

Para enriquecer este indicador, el docente decide diseñar un nuevo indicador. En este caso, calculando el número de palabras por turno que escriben los estudiantes.

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

**Tabla 5.6:** Información sobre las palabras por turnos utilizadas por los estudiantes en el role-play

id	username	words	turns	words per turn
1	student1	117	9	13
2	student2	132	11	12
3	student3	63	9	7
4	student4	140	10	14
5	student5	99	11	9
6	student6	80	10	8
7	student7	72	9	8
8	student8	108	9	12

Para ello utiliza la consulta VWQL 2. Los resultados en este caso puede verse en la tabla 5.6.

```
_____ Consulta VWQL 2 _____  
1 Evidence palabras_turno:  
2   get students  
3   show words, turns.
```

---

A la vista de los resultados, el docente confirma que los estudiantes 3, 6 y 7 (*student3*, *student6* y *student7*) son los que utilizan menos palabras por turno. Son estos tres estudiantes, junto con el estudiante número 5 (*student5*), los únicos que bajan de 10 palabras por turno. Lo que podría ser una justificación para el docente a la hora de evaluar de manera negativa a este estudiante, ya que no sólo no llega a 10 mensajes sino que también es el único que en la consulta anterior está en el límite fijado de dos respuestas cortas.

Como en los casos anteriores, será decisión del docente el uso que se dé a los indicadores.

### Propuesta de indicadores para el ejemplo

El mundo virtual se ha desarrollado para asignaturas de idiomas, por lo que la competencia genérica para la que más se han utilizado los indicadores obtenidos ha sido

para la *habilidad para comunicarse en lengua extranjera* (LAN). No obstante, los indicadores podrían ser utilizados en la evaluación de otras competencias genéricas bajo el criterio del docente que diseña la evaluación.

#### **Lengua extranjera (LAN)**

- *Ritmo (fluidez)*: número de frases escritas por minuto. Según la experiencia desarrollada, un valor alto podría ser un indicador positivo o negativo del desempeño de la competencia. Si el estudiante tiene que contar una historia o expresarse sin restricción, el hecho de que escriba muchas frases por minuto es un indicador positivo de su dominio del idioma. Sin embargo, si el estudiante tiene que enviar a su compañero un mensaje concreto y tiene dificultades para hacerse entender, puede necesitar más de una frase por minuto para hacerse entender, siendo entonces un indicador negativo.
- *Frases por turno*: número de frases escritas por turno. Igual que en el caso anterior, puede interpretarse de distinta forma según el caso. Si el estudiante tiene un buen dominio del idioma y no hay restricción impuesta, puede ser un indicador positivo que escriba muchas frases por turno. Mientras que si necesita muchas frases para transmitir un mensaje concreto, que tenga que escribir muchas frases por turno puede ser un indicador negativo del desempeño de la competencia.
- *Intervenciones de una palabra*: número de frases o intervenciones de una sola palabra escritas por el estudiante. Un abuso del uso de frases de una sola palabra puede ser utilizado como un indicador del bajo nivel de conocimiento de la lengua extranjera.

**Capacidad de aprender y mantenerse al día con el aprendizaje (LLL)** Si un estudiante tiene dificultades para desenvolverse en el idioma, el hecho de que dedique muchos minutos a jugar en el mundo virtual podría ser un indicador de que dicho estudiante está comprometido con su aprendizaje, y quiere mejorar. Sin embargo, si un estudiante tiene malos resultados, y además pasa pocos minutos practicando, sería un indicador negativo de esta competencia.

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

### Resumen

En la tabla 5.7 se muestra el resumen con los indicadores propuestos para la evaluación de competencias genéricas a partir del uso de la herramienta EvalSim y el lenguaje VWQL. Los indicadores para medir la competencia LAM fueron utilizados en la experiencia presentada en [9].

COMPETENCIA	INDICADOR
LAN	Frases escritas por minuto
LAN	Frases escritas por turno
LAN	Abuso del uso de intervenciones de una sola palabra
LLL	Tiempo dedicado a practicar en el mundo virtual

**Tabla 5.7:** Resumen de propuesta de indicadores a partir del uso de EvalSim

### 5.4 Estudios de caso

Entre los cursos 2010-11 y 2014-15 se han realizado diversas experiencias para la evaluación de competencias genéricas utilizando el método DBA y las herramientas que lo implementan. El resumen de herramientas utilizadas, junto con las competencias evaluadas se muestra en la tabla 5.8, mientras que las publicaciones en las que se han presentado estos estudios de caso se detallan en la tabla 5.9. A continuación se describen brevemente cada uno de los estudios de caso.

#### 5.4.1 Estudio de caso CS01

Este estudio se llevó a cabo en la asignatura de *Administración de Sistemas Operativos* durante los cursos 2010-11 y 2011-12, con 38 y 40 estudiantes respectivamente. Esta asignatura tenía lugar en el segundo semestre del tercer (y último) curso de la titulación de Ingeniería Técnica en Informática de Sistemas. Los estudiantes tuvieron que desarrollar varias tareas durante el curso, dos de ellas en un wiki público<sup>1</sup>. La primera consistió en un proyecto de migración de la infraestruc-

---

<sup>1</sup><http://wikis.uca.es/wikiASO>



## 5.4 Estudios de caso

**Tabla 5.8:** Resumen de competencias medidas en cada estudio de caso

Estudio de caso	AUT	COM	CRI	INT	KNO	LAN	LEA	PLA	PRO	QUA	TMW
CS01		X	X		X					X	
CS02				X			X				
CS03			X								
CS04	X			X	X			X	X		X
CS05	X			X	X			X	X		X
CS06						X					

**Tabla 5.9:** Resumen de publicaciones de los estudios de caso

Estudio de caso	Congreso	Revista
CS01	SPDECE 2012 ([12])	CHB ([78])
CS02	TEEM 2013 ([14])	IJEE ([10])
CS04		
CS03		
CS05	SIIE 2014 ([11])	
CS06	TEEM 2015 ([9])	JITR (In process)

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

tura de red de una empresa, mientras que la segunda consistía en documentar un programa Unix para tareas del sistema operativo. Utilizando AMW cada estudiante evaluó 10 contribuciones de sus compañeros al wiki. El docente utilizó AMW y las evaluaciones de los estudiantes para evaluar las competencias TMW, CRI, COM, KNO y QUA.

### 5.4.2 Estudio de caso CS02

El primer estudio de caso con EvalCourse tuvo lugar en la asignatura de *Procesadores de Lenguajes II* del curso 2012-13, con 36 estudiantes matriculados. Esta era una asignatura obligatoria, que tenía lugar en el primer semestre del quinto (y último) curso de la titulación de Ingeniería en Informática. Durante el semestre, los estudiantes tuvieron que trabajar en pequeños equipos de dos o tres miembros. Cada equipo del curso tenía un foro para la comunicación interna. El docente utilizó la primera versión de EvalCourse y SASQL para evaluar las competencias genéricas de INT y LEA.

### 5.4.3 Estudio de caso CS03

El segundo estudio de caso con EvalCourse se desarrolló en la asignatura de *Programación Funcional* del curso 2012-13, en la que había 19 estudiantes matriculados. Esta era una asignatura optativa que tenía lugar en el segundo semestre del quinto (y último) curso de la titulación de Ingeniería en Informática. Los estudiantes tuvieron que trabajar en cuatro talleres a lo largo del curso. En Moodle, un taller es un entregable que, conforme a las instrucciones del docente, puede ser evaluado por otro u otros estudiantes o auto-evaluado. El docente utilizó la primera versión de EvalCourse y SASQL para evaluar la competencia genérica CRI a partir del trabajo desempeñado por los estudiantes en los talleres.

### 5.4.4 Estudio de caso CS04

Para este estudio de caso se cursó una solicitud para la realización de un *Proyecto de Actuación Avalada para la Mejora Docente* en la Universidad de Cádiz durante el curso 2013/14. En estos proyectos tienen cabida aquellos trabajos que supongan

mejoras en un amplio espectro de objetivos relacionados con la docencia en los títulos de la universidad. En este caso, el título de nuestro proyecto fue *Extracción de indicadores objetivos para evaluación del desarrollo de competencias genéricas a partir de registros de actividad del Campus Virtual* (código AAA.14.009). En esta actuación se mejoró la herramienta EvalCourse para que pudiera obtener nuevos indicadores y se utilizó la herramienta en una serie de asignaturas del Grado en Ingeniería Informática.

### 5.4.5 Estudio de caso CS05

Para este estudio de caso se cursó una solicitud para la realización de un *Proyecto de Actuación Avalada para la Mejora Docente* en la Universidad de Cádiz durante el curso 2014/15. En este caso, el título de nuestro proyecto fue *Evaluación de competencias genéricas mediante la extracción de indicadores de los registros de actividad del wiki de Moodle* (código sol-201400047964-tra). En esta actuación se mejoró la herramienta EvalCourse para que pudiera obtener indicadores del wiki de Moodle y se aplicó en varias asignaturas del Grado en Ingeniería Informática y una del Grado en Publicidad y Relaciones Públicas.

### 5.4.6 Estudio de caso CS06

Este estudio de caso se llevó a cabo en un instituto en el curso 2014-15, en la asignatura de *Alemán como lengua extranjera*, equivalente al nivel B1 del marco común de referencia europeo (CEFR, del inglés, Common European Framework of Reference). 5 estudiantes participaron durante sus clases en varias partidas en un mundo virtual. El profesorado utilizó EvalSim y VWQL para obtener indicadores de la participación de los estudiantes en el mundo virtual y aplicarlos a la evaluación de la competencia genérica LAN.

## 5.5 Cuestionarios de evaluación

Para la evaluación de este trabajo de investigación se han desarrollado tres cuestionarios dirigidos a profesionales del ámbito docente y de las TIC siguiendo las

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

directrices marcadas por Oates [75] (capítulo 7). Los cuestionarios y sus resultados pueden verse en los apéndices A, B y C.

### 5.5.1 Objetivo

El objetivo de los cuestionarios es evaluar si los objetivos de esta tesis se han alcanzado. Cabe recordar que los objetivos de esta tesis son los siguientes:

- *O1*: Definir un método que permita al docente obtener de manera automática un conjunto de indicadores de los entornos virtuales de aprendizaje
- *O2*: Definir un DSL que permita a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje

Para valorar la consecución de los objetivos se plantearon las siguientes hipótesis:

- *H1*: El método DBA permite obtener de manera automática indicadores de los entornos virtuales de aprendizaje
- *H2*: El DSL permite a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje

Para evaluar tanto *H1* como *H2* se utilizó el cuestionario C.1. En dicho cuestionario se realizaron diversas preguntas relacionadas con el método DBA y con el DSL a docentes de diferentes ramas. De esta forma no sólo se persigue saber si los docentes dan validez al método o el DSL, sino también que la consideración de alguno de los aspectos de estos es independiente de la rama de cada docente.

Además, para saber si los docentes consideran válidos los indicadores obtenidos mediante el método DBA y la herramienta EvalCourse para evaluar las competencias genéricas se utilizaron los cuestionarios A.1 y B.1. Para ello se planteó en el cuestionario un estudio de caso de ejemplo, preguntándose a los participantes si consideraban los indicadores obtenidos mediante el método DBA y la herramienta EvalCourse válidos para evaluar las competencias genéricas.

Como se puede observar en esta tesis a tenor de los objetivos, lo que se propone es un método y la herramienta (DSL) para diseñar evaluaciones de competencias genéricas. Esto significa que los indicadores que se proponen a lo largo de los ejemplos y los estudios de caso mostrados no son fijos, es decir, que lo que es válido para un docente como indicador de una competencia puede no serlo para otro. Por tanto, un número elevado de docentes que considere válido los indicadores no justificaría en exclusiva la validez de los resultados. Para dar validez a los resultados, además hay que demostrar que el hecho de que un número de docentes lo acepte y otros no debe estar marcado única y exclusivamente por el criterio de cada docente, y no porque sus habilidades tecnológicas le confieran mayor pericia y conocimiento en el manejo del DSL que proporciona los indicadores.

### 5.5.2 Distribución

Los cuestionarios fueron completados por diferentes colectivos de profesionales de la docencia y las TIC. De esta forma se pretendía abordar dos cuestiones importantes. En primer lugar, obtener un número representativo de cuestionarios completados, algo que suele resultar complicado cuando se depende de la buena voluntad y el tiempo de los encuestados. Y en segundo lugar, poder analizar la percepción de la herramienta y el método desde la perspectiva de diferentes colectivos relacionados con el dominio. A continuación se describe cada uno de estos colectivos.

#### **Curso formación PDI**

Durante el curso 2014-15 se impartió a docentes de la Universidad de Cádiz un curso de formación sobre wikis cuyo título fue *Introducción al uso educativo de wikis*. El curso se impartió en los cuatro campus de la Universidad de Cádiz, asistiendo en total unos 30 docentes. Cada edición del curso constó de dos sesiones de 4 horas cada una en las que se explicó cómo trabajar en clase con wikis, se propusieron enfoques para favorecer el desempeño de competencias genéricas de los estudiantes, presentándose el método DBA y la herramienta EvalCourse. Además, se incluyeron actividades prácticas que se desarrollaron en las mismas sesiones.

A los asistentes al curso, que fueron todos docentes de la Universidad de Cádiz, se les envió el cuestionario A.1, siendo 11 los que lo completaron.

## **5. RESULTADOS DE LA INVESTIGACIÓN**

---

### **Taller Aulablog**

Los días 6, 7 y 8 de julio del 2015 se celebró en Ubrique (Cádiz) el X encuentro Aulablog<sup>1</sup>. Aulablog es una reunión de docentes de toda España donde se ponen en común experiencias dentro del ámbito de las tecnologías educativas. En dicho encuentro impartimos un taller sobre wikis en educación. El taller constó de una sesión de 3 horas en la que se explicó cómo trabajar en clase con wikis y se propusieron enfoques para favorecer el desempeño de competencias genéricas de los estudiantes, presentándose el método DBA y la herramienta EvalCourse. Al taller asistieron 20 docentes.

A los asistentes al taller, que fueron docentes de varios niveles educativos, se les envió el cuestionario A.1, siendo 14 los que lo completaron.

### **Actuación avalada**

Durante los cursos 2013-14 y 2014-15 dirigimos dos proyectos de Actuación Avalada para la Mejora Docente dentro de las Convocatorias de Innovación Docente de la Universidad de Cádiz. En dichas experiencias colaboraron varios docentes responsables de varias asignaturas de los Grados en Ingeniería Informática y en Publicidad y Relaciones Públicas. Una vez finalizado cada curso se aplicó EvalCourse y se pusieron los indicadores obtenidos a disposición de los docentes para que lo utilizaran y propusieran mejoras tanto de posibles indicadores a obtener como de mejoras en la herramienta.

A partir de este proyecto se consiguieron 8 cuestionarios completados (cuestionario A.1). Este grupo además completó un cuestionario complementario en el que indicaban para qué competencias genéricas podrían utilizar en sus asignaturas los indicadores proporcionados por EvalCourse (Cuestionario B.1).

### **Wikimedia España**

Se implicó a expertos en trabajo colaborativo en wikis, pero no necesariamente ligados a la docencia. Mediante correo electrónico enviado a los miembros de Wikimedia España, se les explicó el método DBA, la herramienta EvalCourse y se les invita a completar el cuestionario.

---

<sup>1</sup><http://www.aulablog.com/blog/encuentro10>

En este caso se consiguieron 19 cuestionarios completados (cuestionario A.1).

### **Jornadas de Innovación Docente**

En estas jornadas celebradas en el mes de marzo de 2016 se presentaron tanto el método DBA como el DSL que lo implementa en los mundos virtuales (VWQL). Los participantes eran docentes de todas las ramas de la universidad que presencialmente completaron la encuesta.

En este último caso se recopilaron 31 cuestionarios (cuestionario C.1).

### **5.5.3 Resultados del cuestionario del método y el DSL**

El cuestionario C.1 fue completado por 31 docentes en el mes de marzo de 2016 tras asistir a la presentación del método y el DSL en las Jornadas de Innovación Docente. El número de participantes se consideró adecuado en base la guía de Oates, que indica que 30 es el mínimo aceptado para un primer proyecto de investigación.

Antes de completar el cuestionario, los docentes asistieron a una presentación oral en la que se mostró tanto el método DBA como VWQL, el DSL creado para aplicar el método a los mundos virtuales. Además, se mostraron ejemplos de consultas y de resultados asociados a dichas consultas.

### **Perfil de los participantes**

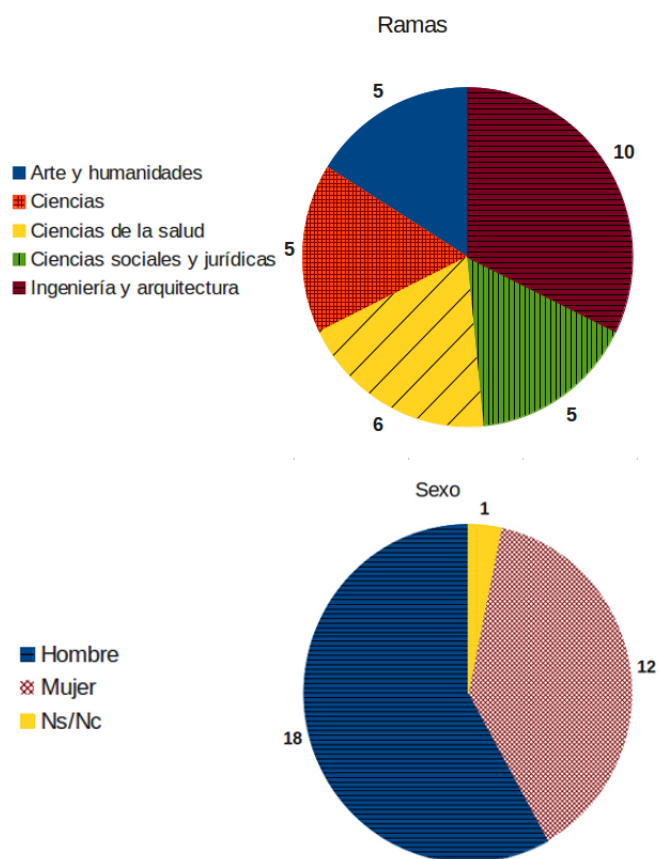
La distribución de los participantes con respecto a la rama que pertenecen y su sexo se muestran en los diagramas de la figura 5.10. Además, en la tabla 5.10 se muestra un resumen de esta información.

### **Evaluación del método DBA**

Para evaluar el método se preguntó a los docentes si consideraban el método DBA adecuado para obtener automáticamente indicadores de los entornos virtuales de aprendizaje. A esta pregunta fueron 26 de los 31 docentes encuestados (84%) los que consideraron el método DBA adecuado (figura 5.11).

En general se puede decir que los docentes consideran el método como válido. Pero también es necesario determinar si existe relación entre la respuesta de cada

## 5. RESULTADOS DE LA INVESTIGACIÓN

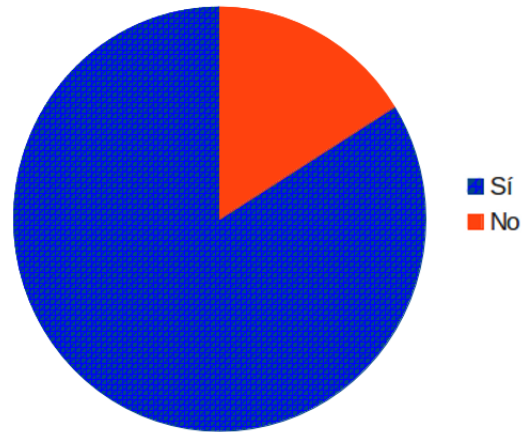


**Figura 5.10:** Distribución de los participantes por rama y sexo

Rama\Sexo	Hombre	Mujer	Ns/Nc	Total
Arte y humanidades	1	4	0	5
Ciencias	3	2	0	5
Ciencias de la salud	6	0	0	6
Ciencias sociales y jurídicas	2	2	1	5
Ingeniería y arquitectura	6	4	0	10

**Tabla 5.10:** Resumen de distribución de participantes por rama y sexo





**Figura 5.11:** Respuestas de los encuestados a la pregunta sobre el método DBA

docente y su rama. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que el método DBA sea o no válido es independiente de la rama de dicho docente*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un p-valor de 0,6142, muy superior al umbral de significación del 0,005.

Para afinar un poco más se decidió además agrupar los docentes según si pertenecen a las ciencias naturales, ciencias de la salud e ingeniería (*EHSE, Earth & Health Sciences and Engineering*) o a ciencias sociales y humanidades (*SSH, Social Sciences and Humanities*). Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que el método DBA sea o no válido es independiente de si dicho docente es de EHSE o de SSH*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un p-valor de 0,2955, superior al umbral de significación del 0,005.

El resumen de las respuestas puede verse en la tabla 5.11.

### **Características del método DBA**

Para continuar con la evaluación del método, se preguntó a los encuestados en qué medida consideran que el método DBA satisface cada una de las características que se esperan del mismo. Estas características son *objetividad, adaptabilidad, sistematicidad, flexibilidad y fiabilidad*. Para su evaluación se preguntó sobre ellos

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

	Rama	Sí	No	Total
SSH	Arte y humanidades	4	1	5
	Ciencias sociales y jurídicas	3	2	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	9	1	10

**Tabla 5.11:** Consideración de idoneidad del método DBA por ramas

en una escala Likert de cinco puntos: totalmente de acuerdo, de acuerdo, ni de acuerdo ni en desacuerdo, en desacuerdo y totalmente en desacuerdo.

En la tabla 5.12 puede verse el porcentaje de votos que obtuvo cada uno de los puntos para cada característica. Las características de la *adaptabilidad*, la *sistematicidad* y la *flexibilidad* son las más aceptadas para el método, ya que sumando los "totalmente de acuerdo" con los "de acuerdo" alcanzan más de la mitad de los votos, esto es, 61, 70 y 61% respectivamente. Mientras que sumando los "en desacuerdo" con los "totalmente en desacuerdo" tienen 16, 10 y 13% respectivamente. Por tanto, se podría decir que la mayoría de los docentes consideran que el método les proporciona estas características.

Para la *objetividad* tenemos un 46% a favor (sumando los "totalmente de acuerdo" con los "de acuerdo") y un 22% en contra (sumando los "en desacuerdo" con los "totalmente en desacuerdo"), mientras que para la *fiabilidad* tenemos un 45% a favor y un 29% en contra, contando ambas características con un 26% de docentes que mantienen una opinión neutral, es decir, ni de acuerdo ni en desacuerdo. Por tanto, podemos decir que aunque con menor porcentaje de docentes a favor, tanto la *objetividad* como la *fiabilidad* son también aceptadas como características del método.

### Evaluación del DSL

Para evaluar el DSL se preguntó a los docentes si consideraban el lenguaje presentado útil para diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje. A esta pregunta también

## 5.5 Cuestionarios de evaluación

Características	Totalmente de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Totalmente en desacuerdo
Objetividad	23%	23%	<b>32%</b>	19%	3%
Adaptabilidad	16%	<b>45%</b>	23%	13%	3%
Sistematicidad	<b>40%</b>	30%	20%	10%	0%
Flexibilidad	19%	<b>42%</b>	26%	13%	0%
Fiabilidad	6%	<b>39%</b>	26%	23%	6%

**Tabla 5.12:** Evaluación de las características del método DBA

fueron 26 de los 31 docentes encuestados (84%) los que consideraron el lenguaje adecuado.

Por tanto se puede decir que los docentes consideran el método como válido. Pero también se quiso determinar si existía relación entre la respuesta de cada docente y su rama. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que el lenguaje sea o no útil es independiente de la rama de dicho docente*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un p-valor de 0,6142, muy superior al umbral de significación del 0,005.

Para afinar un poco más se decidió además agrupar los docentes según si pertenecen a las EHSE o a las SSH. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que el lenguaje sea o no válido es independiente de si dicho docente es de EHSE o de SSH*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar tampoco la hipótesis con un p-valor de 0,2955, superior al umbral de significación del 0,005.

El resumen de las respuestas puede verse en la tabla 5.13.

### Características del DSL

Para continuar con la evaluación del DSL, se preguntó a los encuestados sobre su consideración acerca de en qué medida el DSL satisface cada una de las características que se esperan del mismo. Estas características son *facilidad de aprendizaje*,

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

	Rama	Sí	No	Total
SSH	Arte y humanidades	4	1	5
	Ciencias sociales y jurídicas	3	2	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	9	1	10

**Tabla 5.13:** Consideración de idoneidad del DSL por ramas

*eficacia, ahorra tiempo, escalabilidad, capacidad de reutilización y fiabilidad.* Para su evaluación se preguntó sobre ellos en una escala Likert de cinco puntos: totalmente de acuerdo, de acuerdo, ni de acuerdo ni en desacuerdo, en desacuerdo y totalmente en desacuerdo.

En la tabla 5.14 puede verse el porcentaje de votos que obtuvo cada uno de los puntos para cada característica. En primer lugar cabe destacar que el punto más votado en cuatro de las seis características fue el neutral, es decir, ni de acuerdo ni en desacuerdo. Estas cuatro características son la *fiabilidad* con un 42%, la *facilidad de aprendizaje* con un 39%, la *escalabilidad* con un 35% y el *ahorro de tiempo* con un 32% (en este último caso hay empate con el punto "totalmente de acuerdo").

A pesar de esta "neutralidad", también es importante resaltar que hay un bajo porcentaje de desacuerdos para las características, siendo todos los porcentajes inferiores al 20%, excepto para la *fiabilidad* que es un 23%. Además, para todas las características son más los docentes que están de acuerdo que los que no, superando en cuatro de seis el 50% de votos a favor (*facilidad de aprendizaje* con un 51%, *eficacia* con un 58%, *ahorro de tiempo* con un 55% y *capacidad de reutilización* con un 58%).

### **Evaluación de la sintaxis del DSL (consulta tipo)**

Para evaluar la sintaxis del DSL se mostró una consulta de ejemplo y se preguntó a los docentes si consideraban la consulta entendible. A esta pregunta fueron 24 de los 31 docentes encuestados (77%) los que consideraron entendible la consulta. Por tanto se puede decir que los docentes consideran el método como válido.

## 5.5 Cuestionarios de evaluación

Características	Totalmente de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Totalmente en desacuerdo
Facilidad de aprendizaje	19%	32%	<b>39%</b>	7%	3%
Eficacia	13%	<b>45%</b>	26%	16%	0%
Ahorra tiempo	<b>32%</b>	23%	<b>32%</b>	13%	0%
Escalabilidad	23%	23%	<b>35%</b>	19%	0%
Capacidad de reutilización	26%	<b>32%</b>	29%	6,5%	6,5%
Fiabilidad	6%	29%	<b>42%</b>	23%	0%

**Tabla 5.14:** Evaluación de las características del DSL

Pero también se quiso determinar si existía relación entre la respuesta de cada docente y su rama. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que si la consulta es o no entendible es independiente de la rama de dicho docente*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un p-valor de 0,121, superior al umbral de significación del 0,005.

Para afinar un poco más se decidió además agrupar los docentes según si pertenecen a las EHSE o a las SSH. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que la consulta es o no entendible es independiente de si dicho docente es de EHSE o de SSH*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar tampoco la hipótesis con un p-valor de 0,021, superior al umbral de significación del 0,005.

El resumen de las respuestas puede verse en la tabla 5.15.

### Evaluación de los resultados

Para evaluar los resultados devueltos por el DSL se mostró como ejemplo los resultados que devolvería la consulta de ejemplo y se preguntó a los docentes si consideraban dichos resultados entendibles. A esta pregunta fueron 20 de los 31 docentes

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

	Rama	Sí	No	Total
SSH	Arte y humanidades	2	3	5
	Ciencias sociales y jurídicas	3	2	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	9	1	10

**Tabla 5.15:** Consideración de la consulta como entendible o no por ramas

encuestados (64%) los que consideraron entendible la consulta.

Por tanto se puede decir que los docentes consideran el método como válido. Pero también se quiso determinar si existía relación entre la respuesta de cada docente y su rama. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que si los resultados son o no entendibles es independiente de la rama de dicho docente*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un p-valor de 0,010, superior al umbral de significación del 0,005.

Para afinar un poco más se decidió además agrupar los docentes según si pertenecen a las EHSE o a las SSH. Para ello, se define como hipótesis nula que *la consideración por parte de un docente de que los resultados son o no entendibles es independiente de si dicho docente es de EHSE o de SSH*. Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar tampoco la hipótesis con un p-valor de 0,013, superior al umbral de significación del 0,005.

El resumen de las respuestas puede verse en la tabla 5.16.

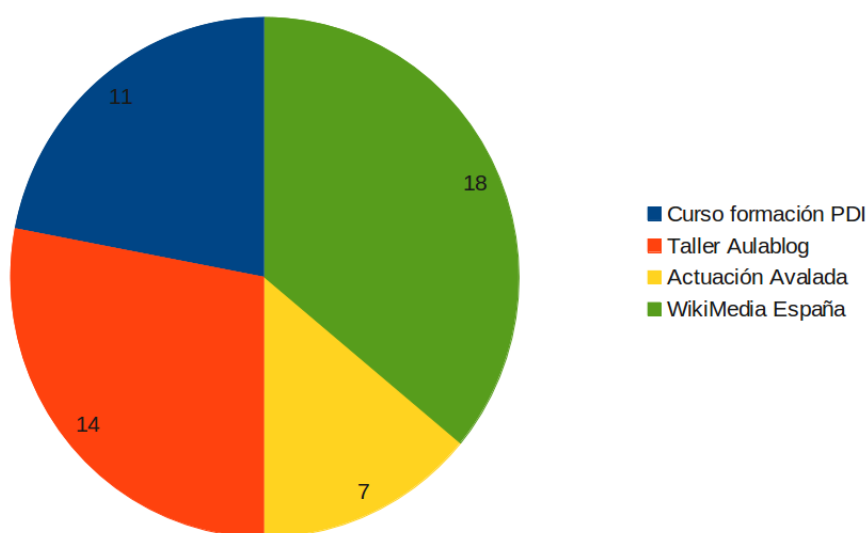
### 5.5.4 Resultados del cuestionario de indicadores

El cuestionario de indicadores (apéndice A) fue enviado en el mes de julio de 2015 a los diferentes colectivos. En este cuestionario se presentaba un estudio de caso de ejemplo en un wiki de Moodle, se mostraban dos consultas escritas en SASQL para la extracción de indicadores aplicables en la evaluación de competencias genéricas y se mostraban algunos de los indicadores y figuras obtenidas mediante la ejecución de dichas consultas en EvalCourse.

## 5.5 Cuestionarios de evaluación

	Rama	Sí	No	Total
SSH	Arte y humanidades	3	2	5
	Ciencias sociales y jurídicas	0	5	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	7	3	10

**Tabla 5.16:** Consideración de los resultados como entendibles o no por ramas



**Figura 5.12:** Distribución de los participantes en el cuestionario

En total se consiguieron 51 cuestionarios completados, cifra que es adecuada a partir de la guía de Oates, que indica que 30 es el mínimo aceptado en una muestra adecuada para un primer proyecto de investigación. La distribución de los participantes con respecto a los grupos a los que pertenecen se muestra en la figura 5.12.

### Perfil de los participantes

El perfil de los participantes gira en torno a dos atributos. Estos atributos son los conocimientos de programación y la relación con la docencia. En la tabla 5.17 se muestra la distribución de los participantes en base a estos atributos.

## 5. RESULTADOS DE LA INVESTIGACIÓN

	Población	Curso formación PDI	Taller Aulablog	Actuación Avalada	Media Wiki Es.	Total
Conocimientos programación	Nulos/ básicos	11	9	0	6	26 (51%)
	Medios/ avanzados	0	5	7	13	25 (49%)
	Totales	11	14	7	19	51
Relación con la docencia	Universitario	11	2	7	1	21 (41%)
	No universitario	0	12	0	0	12 (23%)
	No profesorado	0	0	0	18	18 (35%)

**Tabla 5.17:** Perfil de los participantes que participaron en el cuestionario

Con respecto a los conocimientos técnicos prácticamente hay un empate (26 con conocimientos nulos o básicos y 25 con conocimientos medios y avanzados). Hay que reseñar que fue necesario unir las respuestas de los participantes, ya que las poblaciones intermedias resultaron ser poco significativas y a la hora de aplicar estadísticos se requería un mínimo de población. Por eso los usuarios con conocimiento medio se unieron al grupo de los de conocimiento avanzado y los de conocimiento básico se unieron al grupo de los de conocimiento nulo.

Con respecto a la relación con la docencia vemos que también está bastante repartido (25 docentes universitarios, 12 docentes no universitarios y 35 no docentes). En este caso ocurrió lo mismo. Los docentes no universitarios estaban repartidos entre docentes de infantil, primaria y secundaria. Estos grupos por separado resultaron ser muy poco significativos, por lo que se unieron también en un único grupo a fin de poder aplicar los estadísticos.

A partir de esta distribución de los usuarios se procede a evaluar la idoneidad del método y la herramienta para la evaluación de competencias genéricas en el siguiente apartado.



## 5.5 Cuestionarios de evaluación

RESPUESTA	TMW	PLA	LEA
Sí	26 (51%)	40 (78,4%)	17 (33,3%)
No	25 (49%)	11 (21,6%)	34 (66,7%)

**Tabla 5.18:** Resumen de la validez de cada indicador según los participantes en el cuestionario para evaluar cada competencia genérica

### Evaluación de competencias genéricas

En el cuestionario se pregunta por, según el criterio del participante, la idoneidad de un conjunto de indicadores para evaluar las competencias genéricas TMW, PLA y LEA. El resumen de las respuestas dadas puede verse en la tabla 5.18.

Vemos que la tendencia de las respuestas de los participantes es diferente para cada competencia. Mientras que para evaluar la competencia TMW podríamos decir que hay un empate entre el ‘sí’ y el ‘no’ (26 y 25 respectivamente), en la competencia PLA gana ampliamente el ‘sí’ (40 a 11) y en la competencia LEA gana, aunque no tan ampliamente, el no (17 a 34). Como puede verse las cifras son diferentes para cada competencia, lo que viene a confirmar lo que se viene apostillando en esta tesis desde un principio: el indicador que es válido para un docente para evaluar competencias genéricas puede no serlo para otro. A continuación se aborda la independencia de las respuestas de los participantes en base a su perfil y sus conocimientos técnicos.

### Independencia de las respuestas

Para probar la independencia de las variables que intervienen en el cuestionario se utiliza la prueba  $\chi^2$  de Pearson mediante la presentación de los datos en tablas de contingencia. Para cada competencia genérica se lanzaron dos hipótesis nulas. La primera con el objetivo de demostrar la independencia entre la respuesta de los participantes al uso o no de los indicadores y sus conocimientos de programación. La segunda para demostrar la independencia entre la respuesta y el perfil docente del participante. En total seis hipótesis nulas que fueron las siguientes:

1. Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

competencia **TMW**

2. Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia **PLA**
3. Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia **LEA**
4. El **perfil en relación a la docencia** del individuo es independiente de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia **TMW**
5. El **perfil en relación a la docencia** del individuo es independiente de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia **PLA**
6. El **perfil en relación a la docencia** del individuo es independiente de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia **LEA**

Se construyeron las tablas de contingencia, se calcularon las frecuencias esperadas y se obtuvieron los diferentes valores de  $\chi^2$  para cada hipótesis. Una vez realizados los cálculos las hipótesis nulas 1, 2, 4, 5 y 6 no se rechazaron, concluyéndose que con una significación del 5% los datos son independientes. La hipótesis nula 3 sí quedaba rechazada con ese nivel de significación. Para no rechazarla el nivel de significación debía ser de un 0,1%. El hecho de que la hipótesis nula 3 fuese más difícil de demostrar se debe a que la mayoría de los usuarios que dijo 'sí' a la posibilidad de evaluar la competencia LEA fueron los participantes en el curso de formación de PDI y el taller de Aulablog. En dichas sesiones, con más interactividad que otras actuaciones, se mostraron ejemplos de cómo detectar a un usuario que ejercía una labor de liderazgo en un equipo de trabajo en wiki y esto pudo influir en el hecho de que a dichos participantes les costase menos interpretar en dichos indicadores la posibilidad de utilizarlos para realizar la evaluación que a otros.

El estudio completo puede verse en la sección A.2 del apéndice A.

### **Análisis de las justificaciones**

Los participantes en el cuestionario podían justificar con un comentario el porqué daban el ‘sí’ o el ‘no’ a la posibilidad de utilizar los indicadores a cada competencia. El listado de justificaciones completo puede verse en el apartado A.2.3.3. A continuación se resumen las justificaciones para cada competencia.

**Trabajo en equipo (TMW).** Los participantes que consideran que sí se puede evaluar el TMW argumentan que los indicadores reflejan cómo se han coordinado los estudiantes, la periodicidad de sus contribuciones y sobre todo el volumen de lo aportado al trabajo.

Sin embargo, encontramos participantes que en estos indicadores sólo ven aportaciones individuales, y demandan que para medir la competencia TMW necesitarían otro indicador que mida la interacción de los miembros, si ha habido un reparto de trabajo o si alguno está desempeñando algún rol.

**Planificación y gestión del tiempo (PLA).** En esta competencia el indicador que más defienden los participantes para evaluar la competencia PLA es la periodicidad de las contribuciones. De esta forma consideran que queda justificado si se ha distribuido el trabajo a lo largo del curso y por tanto se ha tenido una buena planificación.

En esta ocasión, los comentarios para no utilizar este indicador van en la línea de la falta de información previa. Es decir, para saber si se han planificado bien, tengo que saber de antemano en qué consistía el trabajo, cómo estaba organizado el curso, si pueden haber existido otros factores que hayan influido, etc.

**Liderazgo (LEA).** Para justificar el liderazgo la mayoría de los participantes se basan en el nivel de participación, es decir, consideran que un estudiante ha ejercido un papel de liderazgo porque ha contribuido al wiki más que sus compañeros. Además, inciden en que el líder dedica al wiki más tiempo que sus compañeros, con contribuciones iniciales que animan a los demás, y con contribuciones finales, para corregir o supervisar el trabajo de estos.

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

**Tabla 5.19:** Cuadro de indicadores considerados para cada competencia

Competencia	Accesos	Foros	Actividades	Wiki	Talleres
TMW	1	2	2	4	0
PLA	5	2	6	5	0
CRI	0	4	2	1	0
PRO	0	5	4	2	1
INT	2	6	0	3	1
AUT	6	3	5	7	1

Los participantes que abogan por el ‘no’ se basan principalmente en que el nivel de participación no implica liderazgo, por lo que descartan totalmente el indicador proporcionado para medir la competencia LEA. Se propone como alternativa opuesta que un indicador de LEA podría ser considerar precisamente qué usuario es el que menos trabaja, siendo necesario complementar este indicador con otro relativo a la comunicación entre los miembros del equipo.

### 5.5.5 Resultados de la encuesta complementaria de indicadores

Además del cuestionario principal, los docentes que participaron en las actuaciones avaladas recibieron un cuestionario en el que valoraban para qué competencias genéricas consideraban útiles los indicadores proporcionados por la herramienta EvalCourse (apéndice B). Esta encuesta también tuvo lugar en julio de 2015, siendo 8 los docentes que participaron en la misma. La tabla 5.19 muestra para cada competencia genérica (filas) cuántos docentes la consideraron evaluable a partir de alguno de los indicadores (columnas).

En este caso, los resultados vienen desglosados por cada actividad del campus virtual. Cabe destacar que la competencia genérica a la que más se pueden aplicar los indicadores de actividad del campus virtual para su evaluación según los docentes que participaron en la encuesta es el trabajo autónomo (AUT), seguida de la planificación y gestión del tiempo (PLA). En la sección B.2 del apéndice B pueden verse al completo las justificaciones de los docentes para considerar los registros

de cada una de las actividades aquí indicadas como indicadores de competencias genéricas.

### 5.5.6 Discusión y análisis

Los resultados de los cuestionarios son muy satisfactorios. En primer lugar, se han evaluado el método DBA y el DSL que implementa el método, con el fin de medir la consecución de las hipótesis planteadas en esta tesis para medir la consecución de los objetivos. En segundo lugar, se han evaluado los resultados obtenidos mediante el DSL, con el fin de valorar su aplicabilidad en la evaluación de diferentes competencias genéricas.

#### Método DBA y DSL

El 84% de los docentes encuestados consideran adecuado el método DBA para obtener indicadores de los entornos virtuales de aprendizaje y consideran el DSL útil para diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de dichos entornos.

A las encuestas se les pasó el test de Fisher para valorar si las respuestas de los docentes son independientes de la rama a la que pertenecen. No se pudo rechazar la hipótesis de independencia de las respuestas para ningún caso, quedando en ambos casos el estadístico muy por encima del valor umbral.

Para afinar aún más en el detalle, se volvió a realizar el test de Fisher tras agrupar a los docentes en dos grupos, según pertenezcan a las SSH (*Social Science and Humanities*) o a las EHSE (*Earth and Health Science and Engineering*). En este caso, tampoco se pudo rechazar la hipótesis de independencia, quedando siempre el valor del estadístico por encima del valor umbral.

Además, el 77% considera entendible la sintaxis utilizada para definir las consultas en el DSL, mientras que el 64 % considera entendibles los resultados obtenidos a partir de la ejecución de dichas consultas.

También en estos casos se aplicó el test de Fisher para valorar la independencia de los resultados, primero con respecto a las ramas de los docentes y segundo con respecto al agrupamiento de dichas ramas según sean de SSH o de EHSE. En todos los casos, el estadístico quedó por encima del umbral, aunque no con una diferencia

## 5. RESULTADOS DE LA INVESTIGACIÓN

---

tan amplia como en los casos anteriores. Sobre todo es llamativo lo que ocurre con los resultados, donde todos los docentes del área de ciencias sociales marcaron como no entendibles los resultados que devolvía el DSL. A pesar de ello, se sigue sin poder rechazar la hipótesis de independencia.

Por tanto, a partir de los resultados obtenidos se acepta la consecución de las hipótesis planteadas en esta tesis:

- *H1*: El método DBA permite obtener de manera automática indicadores de los entornos virtuales de aprendizaje
- *H2*: El DSL permite a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje

Además, se ha preguntado a los encuestados tanto si el método DBA como el DSL han satisfecho las características que se les espera. Más de la mitad de los docentes (entre el 60 y el 70%) considera que el método DBA satisface las características de la *adaptabilidad*, *sistematicidad* y la *flexibilidad*. Las características de la *objetividad* y la *fiabilidad* también tienen una mayor cantidad de votos a favor que en contra, aunque con menor ventaja que las anteriores.

Con respecto a las características del DSL siguen siendo, para todas las características, más los docentes que están a favor que los que están en contra. Sin embargo, en torno un tercio de los encuestados para cada característica se sitúa en la posición neutra. Las características de la *facilidad de aprendizaje*, la *eficacia*, el *ahorro de tiempo* y la *capacidad de reutilización* tiene un porcentaje de docentes a favor de entre el 50 y el 58%. La *escalabilidad* cuenta con 46% a favor y la *fiabilidad* con un 35%.

### Indicadores de competencias genéricas

Aunque en esta tesis se propone un método y una herramienta para la evaluación de competencias genéricas, no es un objetivo de esta tesis establecer un indicador fehaciente para ninguna de ellas. Ese objetivo necesitaría una investigación específica en ciencias sociales y de la educación. En los ejemplos mostrados se proponen

## 5.5 Cuestionarios de evaluación

---

ejemplos de indicadores, y según el criterio de cada participante en el cuestionario, unos son más válidos que otros. Para la competencia TMW prácticamente hay un empate entre el 'sí' (26) y el 'no' (25). Mientras que en la competencia PLA domina el 'sí' (40 a 11), en la competencia LEA domina el 'no' (17 a 34). Las justificaciones de los participantes en muchos casos para dar el 'no' van en la línea de la necesidad de completar el indicador con información previa. Esto podría ser un ejercicio de redefinición de la hipótesis inicial o un rediseño de la evaluación.

En segundo lugar, hay evidencias significativas a favor de que las respuestas de los participantes en la encuestas son independientes tanto del perfil docente como del perfil más o menos técnico de estos. A esta conclusión se llega después haber establecido una serie de hipótesis nulas y haber aplicado la prueba  $\chi^2$  de Pearson. De esta manera se descarta que las propuestas de diseño de evaluación de competencia genérica tengan una orientación demasiado técnica y que por tanto sean los usuario con perfil técnico los únicos que acepten los diseños, o que esté más cercano a un docente de perfil universitario.

El cuestionario complementario utilizado en las Actuaciones Avaladas muestra que los docentes consideraron aplicar los indicadores obtenidos mediante EvalCourse en el campus virtual de Moodle a varias competencias genéricas. Las respuestas de los participantes difieren prácticamente para cada actividad y competencia genérica. Este hecho, al igual que la diferencia de criterio en las respuestas a 'sí' y 'no' a las competencias genéricas del cuestionario principal, respaldan aún más si cabe la propuesta de esta tesis doctoral, es decir, la necesidad de proponer este método a los docentes para que cada uno sea capaz de construir los indicadores que les sean válidos.





# Conclusiones y trabajo futuro

En este capítulo se presentan las conclusiones de este trabajo de investigación, y se comentan los trabajos futuros que se derivan del mismo.

## 6.1 Conclusiones

En esta tesis doctoral se estudia la utilización de indicadores procedentes de los registros de actividades de aprendizaje para evaluar las competencias genéricas de los estudiantes que participan en dichas actividades.

Para alcanzar los objetivos enunciados en el capítulo introductorio se propuso un método de evaluación denominado design-based assessment (DBA) y se desarrollaron dos lenguajes específicos de dominio (DSL). El método DBA permite definir una hipótesis de evaluación mediante el diseño de indicadores a partir de la información de los registros de actividad de los estudiantes en entornos virtuales de aprendizaje, para después analizar los resultados de su aplicación, y si el docente lo considera necesario, redefinir la hipótesis inicial comenzando de nuevo el ciclo. SASQL y VWQL son los DSL que se han desarrollado para facilitar la aplicación

## 6. CONCLUSIONES Y TRABAJO FUTURO

---

del método en cursos virtuales basados en Moodle y en mundos virtuales basados en OpenSim respectivamente.

Para la evaluación tanto del método como de los DSL desarrollados se plantearon las siguientes hipótesis: el método DBA permite obtener de manera automática indicadores de los entornos virtuales de aprendizaje (H1) y los DSL permiten a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje (H2).

Para contrastar dichas hipótesis, el método DBA junto con los DSL se han aplicado y utilizado, en primer lugar, a varias asignaturas del Grado en Ingeniería Informática de la Universidad de Cádiz y en un estudio de caso llevado a cabo en un curso de idiomas con estudiantes de instituto. En segundo lugar, ambos fueron evaluados por docentes de todos los niveles educativos y pertenecientes a todas las ramas de conocimiento.

De los resultados de la evaluación realizada, se deriva que el método DBA permite abordar la evaluación de competencias genéricas, ya que aporta indicadores objetivos del trabajo de los estudiantes en los entornos virtuales que han sido evaluados positivamente en el estudio.

En concreto, hay evidencias a favor de las siguientes conclusiones:

- El método de evaluación DBA es adecuado para diseñar y contrastar estrategias de evaluación a partir de los registros de actividades de los entornos virtuales de aprendizaje. Las características del método DBA más valoradas en el estudio realizado son su adaptabilidad, su sistematicidad y su flexibilidad.
- El DSL facilita la implementación del método DBA. Las características del DSL más valoradas son su facilidad de aprendizaje, su eficacia, su eficiencia y su capacidad de reutilización. Además, tanto las consultas que se pueden definir con el DSL como los resultados que el sistema devuelve fueron fácilmente interpretables por la mayoría de la población encuestada, lo que evidenció que el DSL puede ser utilizado por docentes sin conocimientos de programación.

### 6.1.1 Alcance

El alcance de las conclusiones anteriormente mencionadas se deben considerar cuando concurren situaciones similares a las descritas en este trabajo de investigación. Sin embargo, no se puede confirmar que una situación quede dentro del alcance de las mismas cuando se den circunstancias como las que se describen a continuación:

- Los docentes no incorporen en sus cursos virtuales módulos que generen actividad en los registros de aprendizaje o, si aún teniendo dichos módulos, no fomentan su utilización entre los estudiantes del curso.
- No se disponga de acceso a los registros que contienen la información de la actividad generada por los estudiantes en los entornos virtuales de aprendizaje ni mediante conexión directa ni mediante copia de seguridad. En este caso podría aplicarse un método basado en Linked Open Data [91] para convertir en abiertos los datos alojados en el entorno virtual de aprendizaje.

### 6.1.2 Amenazas a la validez

La principal amenaza a la validez es que el método DBA y los DSL han sido probados y evaluados en una muestra adecuada en tamaño [75] pero limitada. En concreto, la mayoría eran docentes que, aunque no tengan perfil informático, tienen en común su interés por la innovación educativa, las nuevas tecnologías y su aplicación en la mejora de los procesos de aprendizaje. Los foros en los que han conocido tanto el método DBA como los DSL son congresos, seminarios y cursos a los que fueron por propia voluntad. Por tanto, quedaría conocer la experiencia de docentes que a priori no tienen estas inquietudes y/o suelen quedarse al margen de estos eventos.

Por otro lado, también se considera una amenaza a la validez el hecho de que el método DBA, mediante las herramientas implementadas, se ha probado únicamente en dos entornos virtuales de aprendizaje concretos: en el sistema de gestión de aprendizaje Moodle (mediante SASQL) y en mundos virtuales basados en OpenSim (mediante VWQL). Habría que adaptar los DSL para que sean utilizables en

## 6. CONCLUSIONES Y TRABAJO FUTURO

---

el análisis de las interacciones producidas en otros entornos virtuales de aprendizaje tales como plataformas de desarrollo colaborativo, repositorios de contenidos digitales y aplicaciones de realidad aumentada.

### 6.2 Contribuciones

A continuación se indican las aportaciones principales que se han realizado durante la elaboración de este trabajo de investigación:

#### Publicaciones en revistas

Artículo en la revista *Computers in Human Behavior* (2014), titulado “**Scalability of assessments of wiki-based learning experiences in higher education**” [78] (*ISI JCR 2014: Q1, T1*)

En este artículo se presentaron siete estudios de caso de evaluación de wikis en educación superior. Entre esas experiencias destacan dos que hacían uso de herramientas para abordar de manera escalable la evaluación de varias habilidades de los estudiantes a partir del trabajo realizado en wikis de MediaWiki. Se analizaron y compararon las diferentes metodologías y configuraciones de las herramientas utilizadas en el contexto de cada uno de los siete estudios de caso aportados. Una de las herramientas que se utilizaron fue AssessMediaWiki, que mediante procedimientos de autoevaluación, hetero-evaluación y evaluación entre iguales complementó con una visión cualitativa el enfoque cuantitativo aportado por la herramienta StatMediaWiki.

Artículo en la revista *International Journal of Engineering Education* (2015), titulado “**A Domain Specific Language for Online Learning Competence Assessments**” [10] (*ISI JCR 2014: Q3, T3*)

En este artículo se utilizó SASQL para el diseño de evaluaciones a partir de indicadores obtenidos de la interacción de los estudiantes en dos estudios de caso basados en dos cursos virtuales. Estos indicadores

se utilizaron para asistir en la evaluación del desempeño de los estudiantes en las competencias genéricas del liderazgo, las habilidades interpersonales y la habilidad para ser crítico y autocrítico.

Artículo enviado a la revista *Journal of Information Technology Research* (2016) tras ser invitados por una contribución previa en el congreso TEEM 2015, titulado **“Retrieving Objective Indicators from Student Logs in Virtual Worlds”** (Scimago SJR 2014: Q4)

En este artículo se analiza el comportamiento de los estudiantes en un mundo virtual basado en OpenSim mediante el uso de EvalSim y VWQL. Se diseñaron evaluaciones y se obtuvieron indicadores del desempeño de los estudiantes a partir de su interacción en el mundo virtual. Estos indicadores se utilizaron para establecer una comparación más justa del rendimiento de los estudiantes y sacar conclusiones sobre el análisis del proceso de aprendizaje.

### **Contribuciones/Comunicaciones a congresos**

Comunicación presentada en el congreso *IX Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos (SPDECE 2012)*, titulada **“Qualitative assessment of wiki-based learning processes”** [12]

En esta contribución se presentó AssessMediaWiki y se aplicó con el fin de dar soporte a procedimientos de autoevaluación y evaluación entre iguales. Estos procedimientos se utilizaron en la evaluación de las contribuciones realizadas por los estudiantes a un wiki basado en MediaWiki. Los estudiantes evaluaron las contribuciones de sus iguales a partir de los aspectos de una rúbrica sobre el trabajo específico que estos debían realizar, mientras que el docente utilizó dichas evaluaciones para evaluar la capacidad críticas de sus estudiantes.

Comunicación en forma de póster presentada en el *8th European Conference on Technology Enhanced Learning (EC-TEL 2013)*, titulada **“A generative computer language to customize online learning assessments”** [13]

## 6. CONCLUSIONES Y TRABAJO FUTURO

---

En esta contribución se presentaron la herramienta EvalCourse y la sintaxis del lenguaje SASQL junto con un sencillo ejemplo.

Comunicación presentada en el *4th International Workshop on Software Engineering for E-learning (ISELEAR'13)*, dentro del *I International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM 2013)*, titulada “**A generative computer language to customize online learning assessments**” [14]

En esta contribución se presentaron EvalCourse y SASQL, y se aplicaron a un estudio de caso para extraer indicadores de la actividad de los estudiantes del campus virtual de una asignatura. Estos indicadores se utilizaron para asistir en la evaluación del desempeño de estos en las competencias genéricas del liderazgo, la planificación y gestión del tiempo y el trabajo en equipo.

Comunicación presentada en el *Simposio Internacional de Informática Educativa (SIIE 2014)*, titulada “**Domain-driven competence assessment in virtual learning environments. Application to planning and time management skills**” [11]

En esta contribución se compararon dos alternativas para abordar la problemática de la evaluación de competencias genéricas. La primera alternativa se basa en una arquitectura dirigida por modelos en la que se utiliza EvalCourse y SASQL, mientras que la segunda alternativa se basa en Gescompeval\_MD, un servicio Web ReST que permite al docente indicar las competencias que desarrollarán los estudiantes mediante cada actividad. Los resultados de aplicar conjuntamente ambos enfoques mostraron que son complementarios, ya que mientras el servicio web proporciona una retroalimentación mucho más detallada, el enfoque dirigido por modelos es más escalable cuando el número de estudiantes es elevado y hay que evaluar competencias genéricas.

Comunicación presentada en el *6th International Workshop on Software Engineering for E-learning (ISELEAR'15)*, dentro del *III International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM 2015)*, titulada “**A Domain Specific Language to retrieve objective indicators for foreign language learning in virtual worlds**” [9]

En esta contribución se presentó VWQL, así como una experiencia piloto de aplicación en el mundo virtual utilizado en un curso de idiomas. Se diseñaron evaluaciones a partir de indicadores de las interacciones de los estudiantes en el mundo virtual. Estos indicadores se aplicaron en la evaluación de la competencia genérica de la habilidad para comunicarse en lengua extranjera.

### Otras contribuciones

A continuación se describen otras contribuciones que resultan del trabajo desarrollado en esta tesis doctoral y que no se recogen en los apartados anteriores. Estas contribuciones se dividen en dos categorías principales. La primera es la de *contribuciones metodológicas*, categoría en la que se enmarca el método de evaluación propuesto. Mientras que la segunda de las categorías es la de *contribuciones en forma de software*, que engloba las diferentes herramientas desarrolladas para poner en práctica el método.

#### Contribuciones metodológicas

- Método DBA: método de evaluación basado en el diseño de indicadores a partir de la información del registro de actividad de los entornos virtuales de aprendizaje.

#### Contribuciones en forma de software

- Sistema AssessMediaWiki [12]: herramienta web que asiste en procedimientos de evaluación entre iguales y autoevaluación de las contribuciones realizadas por los estudiantes en un wiki basado en MediaWiki.
- Sistema EvalCourse y lenguaje de consultas SASQL [10, 13, 14]: DSL para la aplicación del método DBA en cursos virtuales basados en Moodle.
- Sistema EvalSim y lenguaje de consultas VWQL [9]: DSL para la aplicación del método DBA en mundos virtuales basados en OpenSim.

## 6. CONCLUSIONES Y TRABAJO FUTURO

---

### 6.3 Trabajo futuro

Durante la realización de este trabajo se han superado los retos que se plantearon inicialmente. Sin embargo, se han encontrado algunos problemas o nuevos retos que no han llegado a ser abordados pero que complementarían en muchos casos esta tesis doctoral.

Estos trabajo futuros son:

- Adaptar y ampliar los DSL que se han presentado en esta tesis para que puedan ser utilizados en la obtención de indicadores de otros entornos virtuales de aprendizaje tales como plataformas de desarrollo colaborativo, repositorios de contenidos digitales, y aplicaciones de realidad aumentada.
- Añadir indicadores basados en la aplicación de técnicas de análisis de redes sociales (SNA, del inglés *Social Network Analysis*) a los entornos virtuales de aprendizaje. En un wiki de Moodle utilizado en clase se puso a los estudiantes en parejas para que cada pareja realizase su trabajo en una página del wiki. Una vez finalizada la actividad, aplicando técnicas de SNA se obtuvo el grafo que se muestra en la figura 1. En dicho grafo, el nodo central es el docente. Puede verse cómo al principio el docente creó la sección principal, mientras que los estudiantes ubicaron los enlaces a sus páginas en dicha página principal (arista que une a los nodos periféricos con el nodo central). El resto de aristas que unen normalmente dos nodos periféricos representan los estudiantes que colaboraron juntos en una página. El tamaño de los nodos representa la cantidad de trabajo realizado por cada estudiante. Los indicadores obtenidos mediante este análisis podrían ser aplicados en la evaluación de competencias genéricas relacionadas con el trabajo y la interacción de equipos.
- Integrar los DSL con los propios entornos virtuales de aprendizaje. Los DSL que se han presentado en este trabajo no están integrados en los entornos virtuales, sino que se conectan a ellos para obtener los indicadores. En ocasiones, la configuración de esta conexión puede resultar una tarea compleja





**Figura 6.1:** Grafo resultante de aplicar técnicas de SNA a un wiki de Moodle

para docentes con pocos conocimientos informáticos, por lo que sería interesante poder integrar el DSL en los entornos dentro del perfil de supervisor. Para hacerlo se podría utilizar en el servidor la última versión de Xtext, ya que incorpora XtextServlet para responder a las peticiones de los clientes y así poder implementar el parser del lenguaje para su uso a través de HTTP. En la parte del cliente se podría utilizar un editor web hecho en JavaScript (CodeMirror <sup>1</sup>, Orion <sup>2</sup>, etc.).

- Crear una versión visual del DSL. A pesar de haber tratado de desarrollar un lenguaje sencillo con una sintaxis cercana al dominio de la docencia, puede haber docentes que sigan sin encontrarse cómodos a la hora de programar las consultas. En este caso, una versión visual del DSL podría ser más atractiva para docentes con este perfil. Para desarrollar el DSL visual se podría estudiar su implementación mediante la herramienta Eclipse Sirius <sup>3</sup>.

<sup>1</sup><http://codemirror.net>

<sup>2</sup><https://orionhub.org>

<sup>3</sup><https://eclipse.org/sirius>

## 6. CONCLUSIONES Y TRABAJO FUTURO

---

- Establecer un repositorio que relacione actividades a realizar por los estudiantes en los entornos virtuales de aprendizaje, indicadores que se generan y competencias genéricas que se evaluarían con dichos indicadores. Los docentes que han aplicado la herramienta han ido proponiendo indicadores para evaluar competencias. Si los usuarios del método publican sus hipótesis de evaluación de competencias a partir de indicadores, la comunidad de docentes podría aprovecharse de estos indicadores para ver cómo lo hacen otros docentes y evaluar a sus estudiantes en estas competencias a partir de dichas hipótesis tal cuál están o mediante la redefinición de las mismas.

# Conclusions and future work

This chapter presents the conclusions and the future works derived from this research.

## 7.1 Conclusions

This dissertation studies the use of indicators obtained from virtual learning environments activity registers to assess students' performance in generic skills.

In order to achieve the objectives outlined in the introductory chapter, an evaluation method called Design-Based Assessment (DBA) was proposed and two Domain-Specific Languages (DSL) were developed. The DBA method makes it possible to define a hypothesis evaluation by designing indicators based on virtual learning environments activity registers, to subsequently analyse the results of their application; if the teacher then deems it as necessary, a redefinition of the initial hypothesis may be performed, thus beginning the cycle again. On the other hand, the DSLs developed to facilitate the implementation of the method in both learning management systems based on Moodle and virtual worlds based on OpenSim are called SASQL and VWQL respectively.

The following hypotheses were formulated for the evaluation of the method and the DSLs: the DBA method allows to automatically obtain indicators from virtual

## 7. CONCLUSIONS AND FUTURE WORK

---

learning environments (H1) and the DSLs allow teachers to design and contrast evaluation strategies from virtual learning environments activity registers (H2).

To verify these hypotheses, the DBA method along with the DSLs have been firstly implemented and used in several courses taught in the Degree in Computer Engineering at the University of Cádiz and in a case study conducted in a language course with high school students. Secondly, they were evaluated by teachers belonging to both all educational levels and all branches of knowledge.

From the results obtained from the evaluation, it is concluded that the DBA method can address the assessment of generic skills. The objective indicators of students' interaction in virtual learning environments provided by the DBA method have been positively evaluated in the study.

In particular, there are evidences to support the following conclusions:

- The DBA method is suitable to design and contrast assessment strategies from the virtual learning environments activity registers. The DBA method features which are most valued in the study are its adaptability, its systematicity and its flexibility.
- The DSL facilitates the implementation of the DBA method. The DSL features which are most valued are its ease of learning, its effectiveness, its efficiency and its reusability. In addition, both the queries that can be defined with the DSL and the results that the system returns are easily interpretable by most of the surveyed teachers, thus demonstrating that the DSL can be used by teachers without programming skills.

### 7.1.1 Scope

The scope of the aforementioned conclusions should be considered when situations similar to those described in this research concur. However, it cannot be confirmed that a situation is within their scope when circumstances like the following are occurring:

- Teachers do not include modules in their virtual courses that generate activity in the virtual learning registers or, if even with these modules, teachers do not encourage their students to use them.

- The registers containing information on the activity generated by students in virtual learning environments cannot be accessed neither through direct connection nor through backup. In this case, a method based on Linked Open Data [91] could be applied to turn the activity registered in the virtual learning environment into open data.

### 7.1.2 Threats to validity

The main threat to validity is related to the sample of teachers who participated in the evaluation of the proposals. The DBA method and the DSLs have been tested and evaluated in an adequate sample size [75]. However, the majority of subjects are teachers who, even without a digital profile, share their interest in educational innovation, new technologies and their application in improving learning processes. Forums in which both the DBA method and the DSLs were presented were conferences, seminars and courses that the participants went willingly. Hence, it is still required to know the evaluation of teachers who a priori do not have these concerns and / or tend to stay out of these events.

On the other hand, the fact that the DBA method, using the tools implemented, has been tested only in two specific virtual learning environments is also considered a threat to its validity. These environments are the learning management system Moodle (through SASQL) and virtual worlds based on OpenSim (by VWQL). The DSLs should be adapted to be used in the analysis of interactions produced in other virtual learning environments such as collaborative development platforms, digital content repositories and augmented reality applications.

## 7.2 Contributions

The main contributions made during the preparation of this research are detailed below:

### Journal publications

Journal article in *Computers in Human Behaviour* (2014), entitled “**Scalability of assessments of wiki-based learning experiences in higher education**” [78] (*ISI*

## 7. CONCLUSIONS AND FUTURE WORK

---

*JCR 2014: Q1, T1)*

Seven case studies of assessment in wikis in higher education were presented in this article. Among those experiences, two tools that supported a scalable way to address assessment of various abilities of students from work in MediaWiki wikis stand out. The different configurations and methodologies of the tools were analysed and compared in the context of each of the seven case studies provided. One of the tools used was AssessMediaWiki, which by self-, hetero- and peer-evaluation procedures supplemented the quantitative approach provided by the StatMediaWiki tool with a qualitative view.

Journal article in *International Journal of Engineering Education* (2015), entitled **“A Domain Specific Language for Online Learning Competence Assessments”** [10] (*ISI JCR 2015: Q3, T3*)

This article presented the use of SASQL to design assessments based on indicators derived from the students' interaction in two case studies carried out in two virtual courses. These indicators were used to assist in the assessment of students' performance on the following generic skills: leadership, interpersonal skills and ability to be critical and self-critical.

Journal article submitted to *Journal of Information Technology Research* (2016) invited after presenting in Technological Ecosystem for Enhancing Multiculturality (TEEM 2015) [9], entitled **“Retrieving Objective Indicators from Student Logs in Virtual Worlds”** (SCImago SJR 2015: Q4)

Students' behaviour in a virtual world based on OpenSim was analysed in this paper, using EvalSim and VWQL. Assessments were designed and indicators were obtained from students' interaction in the virtual world. These indicators were used to establish a fairer comparison of students' performance and draw conclusions on the analysis of the learning processes.

### Conference proceedings

Oral communication presented at the *IX Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos (SPDECE 2012)*, entitled **“Qualitative assessment of wiki-based learning processes”** [12]

AssessMediaWiki was introduced and applied in this paper in order to support peer-, hetero- and self-assessment procedures. These procedures were used in students’ contributions to a wiki based on MediaWiki. Students assessed their peer contributions from the aspects of a rubric about the specific work that their peers should perform, whereas the lecturer used the differences in assessments made by the students to assess their critical capacities.

Poster presented at the *8th European Conference on Technology Enhanced Learning (EC-TEL 2013)*, entitled **“A Generative Computer Language to Customize Online Learning Assessments”** [13]

EvalCourse and SASQL syntax were presented in this contribution, along with a simple example.

Oral communication presented at the *4th International Workshop on Software Engineering for E-learning (ISELEAR’13)*, held within *I International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM 2013)*, entitled **“A generative computer language to customize online learning assessments”** [14]

EvalCourse and SASQL were applied to a case study to obtain indicators on students’ activity in a virtual course. These indicators were used to support the assessment of students’ performance on the generic competences of leadership, planning and time management and teamwork.

Oral communication presented at the *International Symposium on Computers in Education (SIIE 2014)*, entitled **“Domain-driven competence assessment in virtual learning environments. Application to planning and time management skills”** [11]

## 7. CONCLUSIONS AND FUTURE WORK

---

This work compares two approaches to deal with the assessment of generic skills. The first approach is based on a model-driven architecture where EvalCourse and SASQL are applied, while the second approach is based on a Web ReST service called Gescompeval\_MD, which allows lecturers to flag the skills that will be developed in each activity. The results of jointly applying both approaches show that they are complementary, since while the web service provides a much more detailed feedback, the model-driven approach is much more scalable when the number of students increases and generic skills have to be assessed.

Oral communication presented at the *6th International Workshop on Software Engineering for E-learning (ISELEAR'15)*, held within *III International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM 2015)*, entitled **“A Domain Specific Language to retrieve objective indicators for foreign language learning in virtual worlds”** [9]

Both VWQL and a pilot experience of its application in a language course were presented in this paper. Several assessments based on indicators were designed from students' interactions in a virtual world. These indicators were applied to the assessment of the 'communication in a second language' generic skill.

### **Other contributions**

This section presents not previously described contributions from the work developed in this thesis. These contributions are divided into two main categories. The first one, called *methodological contributions*, includes the proposed method. The second category, called *software contributions*, encompasses the tools that have been developed to implement the method.



### Methodological contributions

- DBA method: assessment method based on the design of indicators obtained from the information stored in the virtual learning environments activity registers.

### Software contributions

- AssessMediaWiki system [12]: web tool that supports the assessment of contributions made in a wiki based on MediaWiki, through peer-, self- and hetero-assessment procedures.
- EvalCourse system and SASQL query language [10, 13, 14]: DSL oriented to applying the DBA method in Moodle-based learning management systems.
- EvalSim system and VWQL query language [9]: DSL oriented to applying the DBA method in OpenSim-based virtual worlds.

## 7.3 Future work

Although while performing this work, the challenges initially raised have been overcome, some problems and new challenges have been detected. Despite them not having been addressed, it is fair to say that in many cases they would help in complementing this thesis.

These future works are listed below:

- Adaptation and extension of the DSLs that have been presented in this thesis so that they can be used in obtaining indicators of other virtual learning environments such as collaborative development platforms, digital content repositories, and augmented reality applications.
- The addition of indicators based on the application of *social network analysis* (SNA) techniques to virtual learning environments. Students worked in pairs in the wiki of a Moodle-based learning management system, with each pair performing their work in a wiki page. Once the activity was finished, the graph shown in Figure 1 was obtained using SNA techniques. In this graph,

## 7. CONCLUSIONS AND FUTURE WORK

---

the central node represents the teacher. It can be seen how initially the teacher created the main section, while students placed links to their pages in this main page (edge joining the peripheral nodes to the central node). The remaining edges joining two peripheral nodes typically represent students who worked together on a page. The size of the nodes represents the amount of work done by each student. The indicators obtained through this analysis could be applied in the assessment of generic skills related to teamwork and interaction between team members.



**Figura 7.1:** Resulting graphic of applying techniques to SNA Moodle wiki

- The integration of the DSLs into virtual learning environments. Instead of being integrated into virtual learning environments, the DSLs presented in this paper are connected to them in order to obtain the indicators. Sometimes the settings of this connection can represent a complex task for teachers with little computer programming skills. Thus, it would be interesting to integrate DSL environments within the supervisor profile in the virtual learning environment. To do this, the latest version of Xtext could be used in the server,

as it incorporates XtextServlet to respond to customer requests. This way the language parser for use via HTTP could be implemented. On the client side you could use a JavaScript web editor (CodeMirror <sup>1</sup>, Orion <sup>2</sup>, etc.).

- The creation of a visual version of the DSL. Despite having tried to develop a simple language with a syntax close to the domain of teaching, there may be teachers who still feel uncomfortable when coding queries. In this case, a visual version of the DSL could be more attractive, especially to teachers with this profile. The development of the visual DSL could be performed with the Sirius Eclipse tool <sup>3</sup>.
- Creating a repository that maps activities to be performed by students in virtual learning environments, indicators that these activities generate and generic skills that can be assessed with these indicators. The lecturers who have applied the method have proposed different indicators for the assessment of skills. Should they share their skill assessment hypothesis based on indicators, the community of lecturers could take advantage of them to see how other teachers do and assess their students in these skills either by using those assumptions as they are or by redefining them.

---

<sup>1</sup><http://codemirror.net>

<sup>2</sup><https://orionhub.org>

<sup>3</sup><https://eclipse.org/sirius>



APÉNDICE



# Cuestionario de indicadores

Este apéndice contiene el cuestionario utilizado para la evaluación de EvalCourse mediante un estudio de caso de ejemplo que muestra la extracción de indicadores del wiki de Moodle. Además, en un segundo apartado se analizan las respuestas recogidas. Las secciones de este apéndice son:

- Cuestionario (ver sección A.1)
- Resultados (ver sección A.2)

## **A. CUESTIONARIO DE INDICADORES**

---

### **A.1 Cuestionario**

#### **A. Nivel educativo**

Indique el nivel educativo en el que imparte su docencia:

- Infantil
- Primaria
- Secundaria
- Superior
- Otras

#### **B. Conocimientos de programación**

Indique los conocimientos de programación que considera que posee:

- Nulos
- Muy básicos
- Básicos
- Medios/avanzados

#### **C. Consultas de indicadores**

Se muestran dos consultas y las figuras obtenidas mediante EvalCourse. Los tres estudiantes que se mencionan trabajaron en una página del wiki entre las semanas 38 y 50 del curso, momento en que el trabajo debía estar terminado.

C.1 Consulta de indicadores 1

A la vista de la figura A.1 obtenida con la *consulta de indicadores 1*: ¿Sería capaz de valorar si los tres estudiantes han participado activamente en el wiki?

- Sí
- No

\_\_\_\_\_ Consulta de indicadores 1 \_\_\_\_\_

**Evidence** Participacion:  
`get students`  
`show participation`  
`in wiki.`

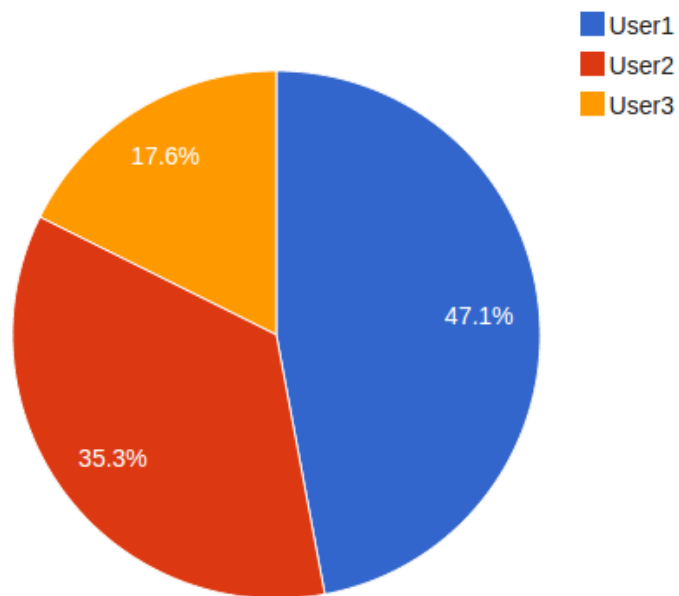


Figura A.1: Participación de los estudiantes en la página del wiki

## A. CUESTIONARIO DE INDICADORES

---

### C.2 Consulta de indicadores 2

A la vista de las figuras A.2 y A.3 obtenidos mediante la *consulta de indicadores 2*: ¿Sería capaz de valorar si los estudiantes contribuyeron al wiki de forma continua?

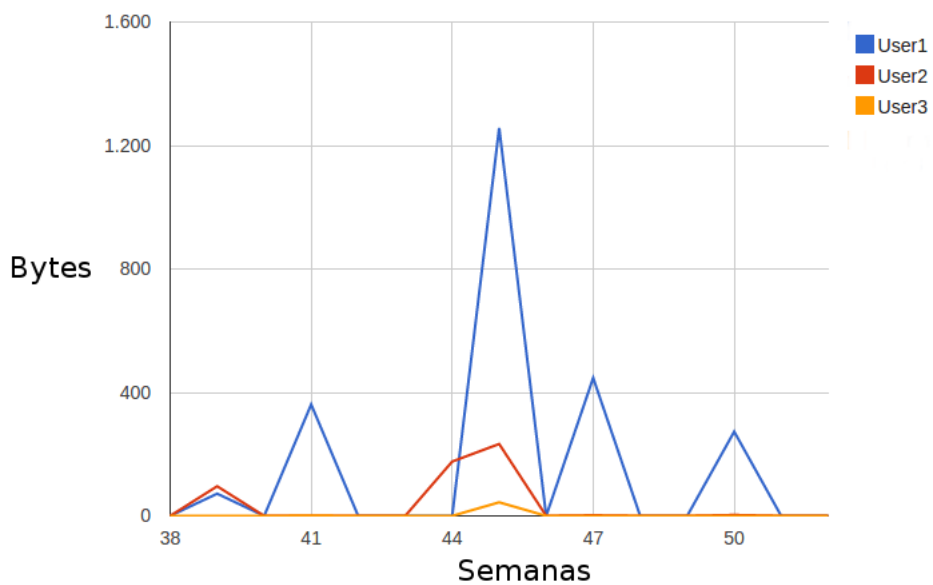
- Sí
- No

\_\_\_\_\_

\_\_\_\_\_ Consulta de indicadores 2 \_\_\_\_\_

**Evidence** Contribuciones:

```
get students
show interaction
in wiki.
```



**Figura A.2:** Contribuciones de los estudiantes a la página del wiki



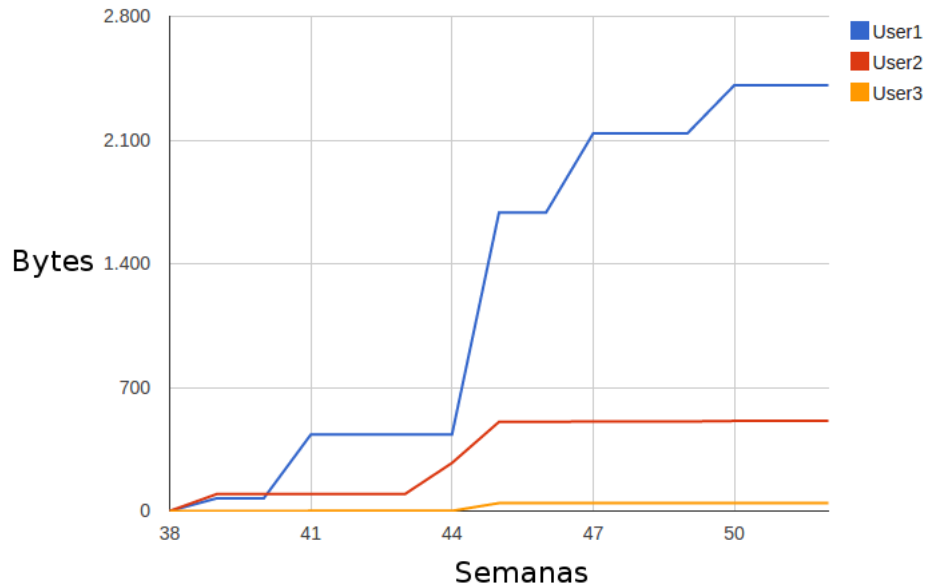


Figura A.3: Evolución del contenido de una página del wiki

## D. Competencias evaluables

Indique para evaluar qué competencias serían de utilidad los gráficos generados anteriormente.

\_\_\_\_\_

Competencia de trabajo en equipo:

• Sí

• No

¿Por qué?

Explique por qué considera que sí o no sería posible evaluar la competencia de trabajo en equipo en el wiki a partir de los gráficos anteriores:

\_\_\_\_\_  
\_\_\_\_\_

Competencia de planificación y gestión del tiempo:

• Sí

## A. CUESTIONARIO DE INDICADORES

---

- No

¿Por qué?

Explique por qué considera que sí o no sería posible evaluar la competencia de planificación y gestión del tiempo en el wiki a partir de los gráficos anteriores:

---

---

Competencia de liderazgo

- Sí
- No

¿Por qué?

Explique por qué considera que sí o no sería posible evaluar la competencia de liderazgo en el wiki a partir de los gráficos anteriores:

---

---

## A.2 Resultados

En esta sección se muestran y analizan en detalle los resultados del cuestionario.

### A.2.1 Distribución

En este cuestionario participaron cuatro tipos de poblaciones. A cada una de ellas se le ha presentado el método DBA y la herramienta EvalCourse en un contexto diferente y con un nivel de detalle diferente, impuesto este por el foro en el que se presentaba. A continuación se presentan cada una de estas poblaciones y el feedback con el que respondieron el cuestionario.

#### **Curso formación PDI**

Profesorado de la Universidad de Cádiz que participó en el curso de formación docente sobre wikis.

- Forma de presentación: Curso presencial.
- Organización: Dos sesiones de 4 horas cada una en la que se explicaba cómo trabajar en clase con wikis, se propusieron enfoques para favorecer el desempeño de competencias genéricas de los estudiantes y se presentaba el método DBA y la herramienta EvalCourse. Además, se incluyeron actividades prácticas que se desarrollaron en las mismas sesiones.
- Número de participantes en el cuestionario: 11

#### **Taller Aulablog**

Profesorado a nivel nacional que asistió al taller sobre wikis en educación que se presentó en el encuentro de profesorado Aulablog 2015.

- Forma de presentación: Taller presencial.
- Organización: Una sesión de 3 horas en la que se explicó cómo trabajar en clase con wikis, se proponían enfoques para favorecer el desempeño de competencias genéricas de los estudiantes y se presentaba el método DBA y la herramienta EvalCourse.

## A. CUESTIONARIO DE INDICADORES

---

- Número de participantes en el cuestionario: 14

### **Actuación avalada**

Profesorado de la Universidad de Cádiz que participó en la actuación avalada sobre el uso de wikis.

- Forma de presentación: Entrevista personal.
- Organización: Se explicó a los participantes el método DBA y se les invitó a utilizar actividades en su LMS (basado en Moodle) que favoreciesen la actividad de los estudiantes en el curso para aplicar después EvalCourse. Se hizo especial hincapié en la inclusión de wikis.
- Número de participantes en el cuestionario: 7

### **Wikimedia**

Miembros de Wikimedia España que fueron invitados y aceptaron participar en el cuestionario.

- Forma de presentación: Correo electrónico.
- Organización: Mediante correo electrónico enviado a los miembros de Wikimedia España, se les explica el método DBA, la herramienta EvalCourse y se les invita a completar el cuestionario.
- Número de participantes en el cuestionario: 19

En total son 51 cuestionarios completados, repartidos tal y cómo se resume en la tabla A.1.

### **Participantes**

Como se muestra en la tabla A.2, los participantes en el cuestionario cuentan con diferentes perfiles. Por un lado nos encontramos con profesores tanto universitarios (41%), como no universitario (23%), es decir, de infantil, primaria o secundaria. Por otro lado han participado en el cuestionario profesionales que no son profesores

POBLACIÓN	CUESTIONARIOS	PORCENTAJE
Curso innovación docente	11	21,57%
Taller Aulablog	14	27,45%
Actuación avalada	7	13,73%
MediaWiki	19	37,25%

**Tabla A.1:** Resumen de participantes en el cuestionario

(35%), pero que tiene un profundo conocimiento en el uso de wikis, ya que son miembros de Wikimedia España.

Además, los conocimientos de programación no deben ser un requisito de los usuarios que utilicen el método DBA y EvalCourse, por lo que se ha tratado de contar con usuarios con un nivel de programación medio o alto (49%), como con conocimientos de programación nulos o básicos (51%).

### A.2.2 Indicadores de participación y de contribución

En este cuestionario se proponen dos consultas escritas en el lenguaje SASQL, los resultados de ejecutarlas en EvalCourse y se pregunta, en primer lugar, si se identifican los resultados de cada consulta como posibles los indicadores de la participación y de la contribución de cada estudiante al wiki.

El objetivo de esta primera pregunta de cada consulta es conocer si el participante identifica los resultados devueltos por EvalCourse como indicadores de la actividad de los estudiantes referidos a la participación y a la contribución de estos al wiki. En la tabla A.3 puede verse cómo de los 51 encuestados la mayoría aceptan los resultados como indicadores de participación (82%) o contribución (88%).

A partir de este punto se pasa a la evaluación de competencias genéricas, donde los participantes indicarán si consideran estos indicadores para la evaluación de competencias genéricas.

## A. CUESTIONARIO DE INDICADORES

---

	Población	Curso formación PDI	Taller Aulablog	Actuación Avalada	Media Wiki Es.	Total
Conocimientos programación	Nulos/ básicos	11	9	0	6	26 (51%)
	Medios/ avanzados	0	5	7	13	25 (49%)
	Totales	11	14	7	19	51
Relación con la docencia	Universitario	11	2	7	1	21 (41%)
	No universi- tario	0	12	0	0	12 (23%)
	No profesorado	0	0	0	18	18 (35%)

**Tabla A.2:** Perfil de los participantes en el cuestionario

Respuesta	Indicador de participación	Indicador de contribución
Sí	42 (82%)	45 (88%)
No	9 (18%)	6 (22%)

**Tabla A.3:** Respuestas dadas a la consideración como indicadores de actividad

Respuesta	Trabajo en equipo	Planificación y gestión del tiempo	Liderazgo
Sí	26 (51%)	40 (78,4%)	17 (33,3%)
No	25 (49%)	11 (21,6%)	34 (66,7%)

**Tabla A.4:** Resumen de la validez dada por los participantes a los indicadores para evaluar competencias genéricas

### A.2.3 Caso de evaluación de competencias genéricas

Los participantes deberán responder a diferentes preguntas sobre si utilizarían los indicadores obtenidos para evaluar tres competencias genéricas de los estudiantes:

- Trabajo en equipo
- Planificación y gestión del tiempo
- Liderazgo

El resumen de las respuestas dadas puede verse en la tabla A.4. En las subsecciones siguientes se describirán en detalle las respuestas para cada uno de los grupos de participantes.

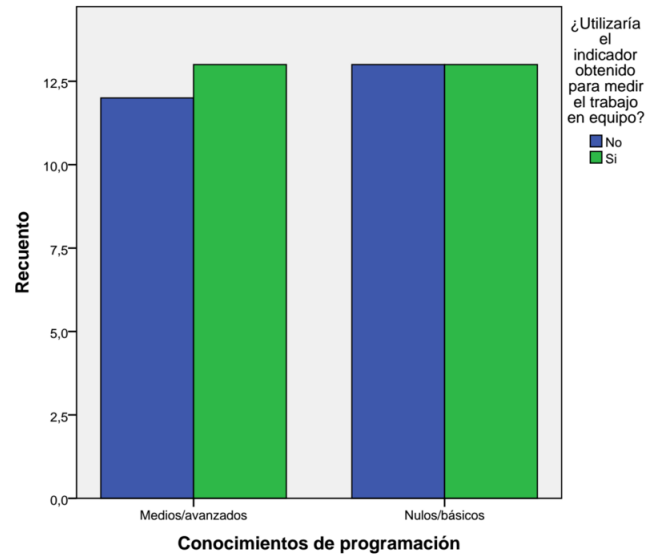
#### A.2.3.1 Conocimientos de programación

¿Que los indicadores de competencias genéricas sean o no válidos para el usuario de EvalCourse es independiente de sus conocimientos de programación? Para responder a esta pregunta se analizan las respuestas dadas a cada una de las tres competencias genéricas con respecto a los conocimientos de programación, y se parte de la hipótesis nula de que ambos valores son independientes.

#### Trabajo en equipo

En la figura A.4 se pueden ver las respuestas de los participantes a si utilizarían o no el indicador proporcionado para evaluar el trabajo en equipo de los usuarios del wiki en base a sus conocimientos de programación.

## A. CUESTIONARIO DE INDICADORES



**Figura A.4:** Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el trabajo en equipo

Para demostrar la independencia entre los conocimientos de programación y la validez que da cada participante al indicador para evaluar dicha competencia vamos a definir la siguiente hipótesis nula:

*$H_0$ : Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **trabajo en equipo***

Para contrastar la hipótesis nula se utilizará la prueba de chi-cuadrado. Se construye la tabla de contingencia y se calculan las frecuencias esperadas, obteniéndose el siguiente valor de chi-cuadrado ( $\chi_0^2$ ):

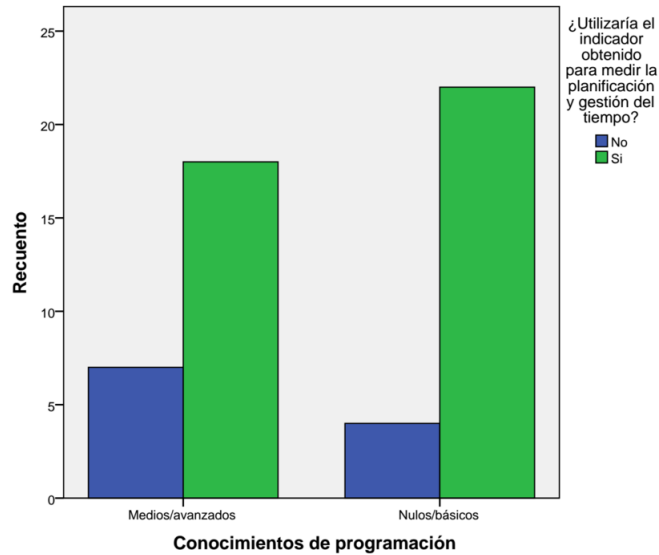
$$\chi_0^2 = 0,02$$

Para contrastar la hipótesis se utiliza un nivel de significación del 5%, siendo este valor 3,84.

$$\chi_{\alpha(r-1)(c-1)}^2 = \chi_{0,05(2-1)(2-1)}^2 = 3,84$$

Como 0,02 es menor que 3,84 no se rechaza la hipótesis nula, y se concluye que con una significación del 5% los datos son independientes.





**Figura A.5:** Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar la planificación y gestión de tiempo

$$\chi_0^2 \leq \chi_{0,05(2-1)(2-1)}^2 \Rightarrow 0,02 \leq 3,84$$

### Planificación y gestión del tiempo

En la figura A.5 se pueden ver las respuestas de los participantes a si utilizarían o no el indicador proporcionado para evaluar la planificación y gestión del tiempo de los usuarios del wiki en base a sus conocimientos de programación.

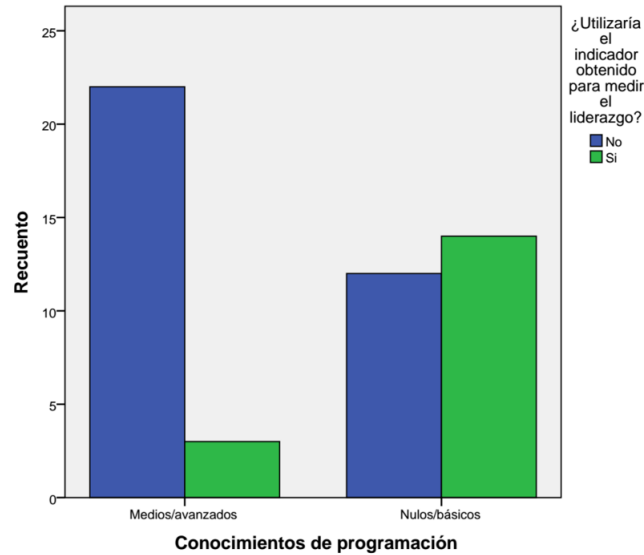
Para demostrar la independencia entre los conocimientos de programación y la validez que da el participantes al indicador para evaluar dicha competencia vamos a definir la siguiente hipótesis nula:

*H<sub>0</sub>: Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **planificación y gestión del tiempo***

Para contrastar la hipótesis nula se utilizará la prueba de chi-cuadrado. Se construye la tabla de contingencia y se calculan las frecuencias esperadas, obteniéndose el siguiente valor de chi-cuadrado ( $\chi_0^2$ ):

## A. CUESTIONARIO DE INDICADORES

---



**Figura A.6:** Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el liderazgo

$$\chi_0^2 = 1,199$$

Para contrastar la hipótesis se utiliza un nivel de significación del 5%, siendo este valor 3,84.

$$\chi_{\alpha(r-1)(c-1)}^2 = \chi_{0,05(2-1)(2-1)}^2 = 3,84$$

Como 1,199 es menor que 3,84 no se rechaza la hipótesis nula, y se concluye que con una significación del 5% los datos son independientes.

$$\chi_0^2 \leq \chi_{0,05(2-1)(2-1)}^2 \Rightarrow 1,199 \leq 3,84$$

### Liderazgo

En la figura A.6 se pueden ver las respuestas de los participantes a si utilizarían o no el indicador proporcionado para evaluar el liderazgo de los usuarios del wiki en base a sus conocimientos de programación.

Para demostrar la independencia entre los conocimientos de programación y la validez que da el participante al indicador para evaluar dicha competencia vamos a definir la siguiente hipótesis nula:

*$H_0$ : Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **liderazgo***

Para contrastar la hipótesis nula se utilizará la prueba de chi-cuadrado. Se construye la tabla de contingencia y se calculan las frecuencias esperadas, obteniéndose el siguiente valor de chi-cuadrado ( $\chi_0^2$ ):

$$\chi_0^2 = 10,043$$

Para contrastar la hipótesis se utiliza un nivel de significación del 5%, siendo este valor 3,84.

$$\chi_{\alpha(r-1)(c-1)}^2 = \chi_{0,05(2-1)(2-1)}^2 = 3,84$$

Como 10,043 es mayor que 3,84 se rechaza la hipótesis nula, y se concluye que con una significación del 5% no se puede asegurar que los datos sean independientes.

$$\chi_0^2 > \chi_{0,05(2-1)(2-1)}^2 \Rightarrow 10,043 > 3,84$$

Para no rechazar la hipótesis nula debería tomarse un nivel de significación del 0,1%, siendo este valor 10,827.

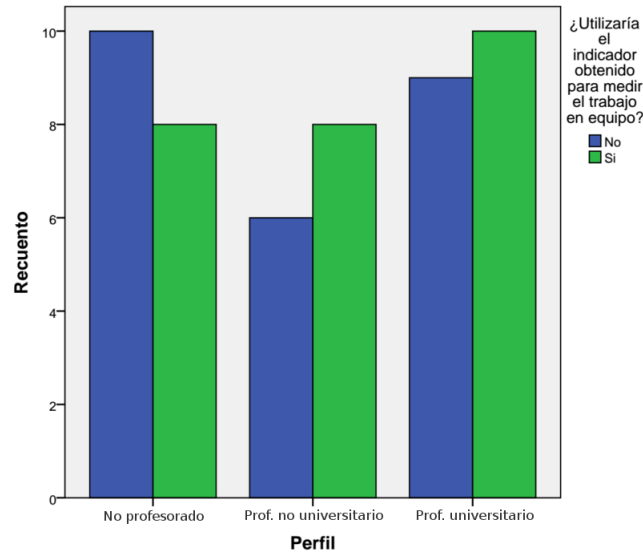
$$\chi_0^2 \leq \chi_{0,001(2-1)(2-1)}^2 \Rightarrow 10,043 \leq 10,827$$

### A.2.3.2 Perfil

#### Trabajo en equipo

En la figura A.7 se pueden ver las respuestas de los participantes a si utilizarían o no el indicador proporcionado para evaluar el trabajo en equipo de los usuarios del wiki en base a su perfil.

## A. CUESTIONARIO DE INDICADORES



**Figura A.7:** Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el trabajo en equipo

Para demostrar la independencia entre el perfil y la validez que da el participantes al indicador para evaluar dicha competencia vamos a definir la siguiente hipótesis nula:

*$H_0$ : El **perfil** del individuo es independiente de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **trabajo en equipo***

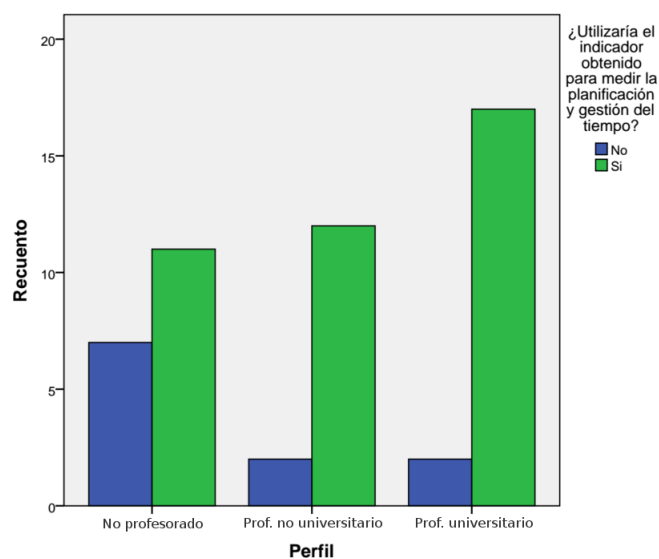
Para contrastar la hipótesis nula se utilizará la prueba de chi-cuadrado. Se construye la tabla de contingencia y se calculan las frecuencias esperadas, obteniéndose el siguiente valor de chi-cuadrado ( $\chi_0^2$ ):

$$\chi_0^2 = 0,541$$

Para contrastar la hipótesis se utiliza un nivel de significación del 5%, siendo este valor 5,991.

$$\chi_{\alpha(r-1)(c-1)}^2 = \chi_{0,05(2-1)(3-1)}^2 = 5,991$$

Como 0,541 es menor que 5,991 no se rechaza la hipótesis nula, y se concluye que con una significación del 5% los datos son independientes.



**Figura A.8:** Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar la planificación y gestión de tiempo

$$\chi_0^2 \leq \chi_{0,05(2-1)(3-1)}^2 \Rightarrow 0,541 \leq 5,991$$

### Planificación y gestión del tiempo

En la figura A.8 se pueden ver las respuestas de los participantes a si utilizarían o no el indicador proporcionado para evaluar la planificación y gestión del tiempo de los usuarios del wiki en base a su perfil.

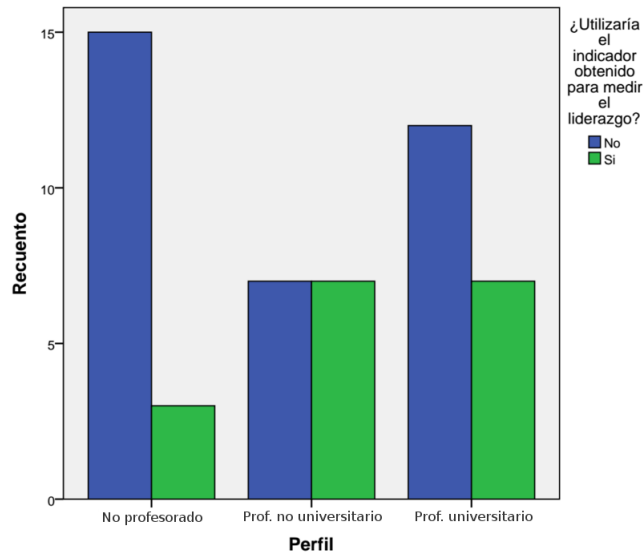
Para demostrar la independencia entre los conocimientos de programación y la validez que da el participante al indicador para evaluar dicha competencia vamos a definir la siguiente hipótesis nula:

*$H_0$ : El **perfil** del individuo es independiente de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **planificación y gestión del tiempo***

Para contrastar la hipótesis nula se utilizará la prueba de chi-cuadrado. Se construye la tabla de contingencia y se calculan las frecuencias esperadas, obteniéndose el siguiente valor de chi-cuadrado ( $\chi_0^2$ ):

## A. CUESTIONARIO DE INDICADORES

---



**Figura A.9:** Respuestas de los encuestados a la pregunta sobre si utilizarían el indicador para evaluar el liderazgo

$$\chi_0^2 = 5,001$$

Para contrastar la hipótesis se utiliza un nivel de significación del 5%, siendo este valor 5,991.

$$\chi_{\alpha(r-1)(c-1)}^2 = \chi_{0,05(2-1)(3-1)}^2 = 5,991$$

Como 5,001 es menor que 5,991 no se rechaza la hipótesis nula, y se concluye que con una significación del 5% los datos son independientes.

$$\chi_0^2 \leq \chi_{0,05(2-1)(2-1)}^2 \Rightarrow 5,001 \leq 5,991$$

### Liderazgo

En la figura A.9 se pueden ver las respuestas de los participantes a si utilizarían o no el indicador proporcionado para evaluar el liderazgo de los usuarios del wiki en base a su perfil.

Para demostrar la independencia entre el perfil y la validez que da el participante al indicador para evaluar dicha competencia vamos a definir la siguiente hipótesis nula:

*H<sub>0</sub>: El **perfil** del individuo es independiente de que el individuo considere que los son válidos los indicadores extraídos para medir la competencia de liderazgo*

Para contrastar la hipótesis nula se utilizará la prueba de chi-cuadrado. Se construye la tabla de contingencia y se calculan las frecuencias esperadas, obteniéndose el siguiente valor de chi-cuadrado ( $\chi_0^2$ ):

$$\chi_0^2 = 4,105$$

Para contrastar la hipótesis se utiliza un nivel de significación del 5%, siendo este valor 5,991.

$$\chi_{\alpha(r-1)(c-1)}^2 = \chi_{0,05(2-1)(3-1)}^2 = 5,991$$

Como 4,105 es menor que 5,991 no se rechaza la hipótesis nula, y se concluye que con una significación del 5% los datos son independientes.

$$\chi_0^2 \leq \chi_{0,05(2-1)(3-1)}^2 \Rightarrow 4,105 \leq 5,991$$

### A.2.3.3 Comentarios

Los participantes en el cuestionario tenían la posibilidad de justificar sus respuestas. A continuación se resume para cada competencia las justificaciones dadas por estos para el sí o para el no:

#### **Trabajo en equipo**

Justificaciones para el sí:

- “Porque las aportaciones de cada uno son muy dispares, demostrando, al menos, poca coordinación entre ellos”

## A. CUESTIONARIO DE INDICADORES

---

- “Pienso que sería posible, ya que podemos comprobar el volumen de aportación de cada uno. Si más o menos los tres han aportado lo mismo, podremos decir que más o menos han trabajado en equipo”
- “Puede medirse el trabajo de cada uno por separado. Sin embargo, no sabemos si son entradas modificadas o cantidad de texto. Quizá el estudiante 3 estuvo más horas trabajando o escribió más texto. Sería interesante medir la cantidad de texto modificado”
- “Si se conoce quién integra los grupos es posible (aunque tal vez no muy fiable) evaluar la correlación entre las curvas de los distintos integrantes y en consecuencia el trabajo en equipo”
- “Porque se percibe el grado de participación de cada integrante”
- “Vemos cuánto ha aportado cada uno y en qué momentos”
- “Saber si el alumno trabajó por su cuenta o realizó un esfuerzo para debatir y buscar consenso en el contenido con sus compañeros”
- “Sería posible observando la proximidad de las gráficas y los bytes que alcanzan, aun en un distinto espacio temporal. Aunque podría darse el caso de que dos aportaciones con el mismo peso en bytes no se correspondan con el tiempo y esfuerzo dedicado, dado que todas las aportaciones no tienen el mismo grado de dificultad”
- “Puede dar información sobre la actividad de cada usuario con respecto al tráfico de datos que ha manejado”
- “Trabajo en equipo = wiki”
- “Se indican las coincidencias temporales y los volúmenes de aportación de información al producto final”
- “Se registra la participación individual de cada alumno al resultado conjunto de grupo”



- “A la vista de la gráfica, suponiendo que se trate de un equipo de trabajo, puede valorarse el grado de implicación, constancia y aportación de cada miembro”
- “Pienso que evaluar la competencia del trabajo en equipo implica evaluar tanto el resultado como el proceso de trabajo. Esta herramienta se centra en el resultado y evidencia el momento en el que se realiza cada una de las aportaciones. Sin embargo, a partir de dichos gráficos, no podríamos derivar aspectos ligados al proceso de trabajo como la capacidad de sus miembros para la planificación, coordinación, liderazgo, resolución de conflictos, etc.”
- “Si los estudiantes verdaderamente han trabajado en equipo los gráficos anteriores deben de reflejar esa contribución colectiva, bien reflejada en tiempos distintos de trabajo más intenso, bien reflejando trabajo en los mismos tiempos, dependiendo de la naturaleza del trabajo encargado por el profesor y la naturaleza de la contribución/colaboración esperada”
- “El gráfico facilita tanto información sobre el número de entradas de cada usuario como el número de aportaciones de cada uno”
- “Se observa que uno de los usuarios ha realizado de manera periódica aportaciones considerables. Los otros dos han participado con menos peso en las mismas fechas. Se puede pensar que el ejecutor principal de la wiki es el 1 y los otros dos aportan ideas ”
- “El trabajo en equipo supone interacción entre los participantes. No me veo capaz de discernir con precisión el grado de interacción con la información proporcionada, ya que quizás distribuyeron previamente el trabajo a realizar, pero por la distribución de trabajo no se ve mucha interacción entre ellos”
- “En los gráficos se puede observar la carga de trabajo de cada miembro del grupo”

Justificaciones para el no:

- “No se puede establecer una relación directa entre la participación individual y la interacción con otros ”

## A. CUESTIONARIO DE INDICADORES

---

- “La mera aportación individual de cada estudiante no parece indicar el trabajo en equipo, ya que pudieron haber realizado las ediciones de forma coordinada o como “lobos solitarios”.”
- “Porque la cantidad de información aportada o el momento en que se haya incluido no tienen que ver con esta competencia, y tampoco la coincidencia de ediciones en el tiempo (que puede ser eso, coincidencia, o bien justo lo contrario a saber trabajar en equipo, es decir, conflictos)”
- “No sabemos en qué espacio de nombres se aportaron esos bytes”
- “Porque no hay nada que indique interacción, sino acción”
- “No se ve sobre qué páginas han interactuado ni cómo lo han hecho”
- “El volumen de aportación no es indicativo del trabajo en equipo porque no hay datos de la calidad de la información aportada. Sí es indicativo de un uso no equitativo. Y aunque la calidad aportada fuera similar, el trabajo en equipo depende del equipo de trabajo. No hay una única forma de trabajar en equipo, por lo que el gráfico no es indicativo”
- “No hay datos sobre el tema”
- “Es necesario conocer qué tipo de trabajo y la composición y roles en el grupo”
- “Se puede observar las aportaciones individuales de cada uno pero no el trabajo colaborativo entre ellos que sería lo importante del trabajo en equipo.”
- “Porque cada alumno podría haber trabajado en áreas totalmente distintas, haciendo un trabajo individual”
- “De los gráficos anteriores no se puede inferir qué cantidad de trabajo han realizado los alumnos en equipo; es posible que cada uno haya realizado sus tareas individualmente por separado”

- “Se puede dar la circunstancia de que los tres trabajen en equipo y sólo uno transcribe las aportaciones, por tanto en el gráfico sólo aparecería la aportación de uno de ellos”
- “Porque sería las aportaciones de los alumnos de forma individualizada. Creo que podría haber otros indicadores para el trabajo en equipo. Las anotaciones entre alumnos u otro indicador parecido”
- “Los gráficos recogen las aportaciones de los alumno individualmente”
- “Por que conoces en qué momento entran y suben cosas, pero no si están trabajando de forma conjunta”

### **Planificación y gestión del tiempo**

Justificaciones para el sí:

- “El usuario 1 realiza contribuciones periódicas todo lo contrario que el usuario 3”
- “Porque se muestra una periodización del trabajo de cada uno”
- “Estudiando la distribución a través del tiempo de las ediciones se puede uno hacer una idea bastante clara de la planificación del tiempo”
- “Porque se aprecia fácilmente la distribución de las contribuciones en el tiempo”
- “Podemos ver si se han planificado para trabajar durante el período estipulado de forma continua o si han tenido picos de trabajo puntuales justo antes de entregar”
- “Puede verse si se ha hecho todo al final o al principio de la temporización”
- “La herramienta permite identificar si el estudiante trabajó por picos o de manera continua, de lo que puede inferirse una planificación (o ausencia de ella) anterior”
- “Porque es posible observar las variaciones en la actividad”

## A. CUESTIONARIO DE INDICADORES

---

- “Vemos cuánto ha aportado cada uno y en qué momentos”
- “Sería posible observando la línea temporal de aportaciones al wiki, comprobando si la carga de trabajo se ha distribuido de forma homogénea a lo largo del tiempo”
- “El gráfico refleja la planificación y gestión del tiempo de cada alumno”
- “Se indican cronológicamente los accesos al objeto del trabajo y los volúmenes aportados”
- “Con las gráficas podemos observar en cada semana el trabajo que ha aportado cada estudiante y se observa el estudiante que ha tenido un trabajo continuado al igual que el que ha realizado su máxima aportación en un periodo concreto”
- “A lo largo del eje temporal se puede valorar la administración del tiempo de trabajo de cada alumno”
- “Puede valorarse si el trabajo ha sido periódico o si se trata de un trabajo hecho en un momento o día determinado”
- “Porque podemos ver cuándo le han dedicado tiempo al trabajo. No se ven las horas, pero podríamos estimarlas a partir del contenido añadido”
- “Por el mismo motivo que en el caso anterior. Muestra cómo el estudiante ha empleado el tiempo en su wiki, y se puede valorar si es así o no como debía haberlo hecho”
- “El gráfico parece indicar que mientras el usuario 1 realiza su actividad en el wiki de forma constante y durante un tiempo más extendido, no ocurre lo mismo con los usuarios 1 y 2, que empiezan a trabajar mucho más tarde y con menos intensidad”
- “Según las gráficas si se observa una simultaneidad en las aportaciones. Por lo que se puede suponer que entre ellos existió una coordinación y planificación del trabajo”

- “En efecto, los gráficos muestran la cantidad de tiempo que los alumnos invierten en sus tareas a lo largo de un período determinado; a partir de aquí podríamos evaluar su capacidad de planificación y gestión del tiempo.”
- “Bueno el 1 ha ido haciendo aportaciones regulares y el 3 sólo al final. Pero puede ser que el reparto que se hicieron del trabajo era precisamente que el 1 buscara la información y el 3 desarrollara conclusiones. Pero bueno, entonces tendría que haber actuado al final y no en el punto intermedio”
- “Se observa claramente la distribución en el tiempo del trabajo realizado”
- “Es fácil comprobarlo viendo la frecuencia de aportaciones, momento, etc”
- “Podríamos comprobar que cada alumno ha estado desarrollando actividades de Wiki a lo largo del tiempo y en qué momento ha trabajado en la actividad”
- “Los gráficos muestran los bytes que cargan los alumnos por semana”
- “Sabemos en qué momento entran, por tanto, si concentran el trabajo o lo hacen de forma continua a lo largo del tiempo establecido.”

Justificaciones para el no:

- “Los gráficos no muestran la gestión del tiempo, es decir, el tiempo dedicado por cada estudiante. La competencia de planificación me resulta un término difícil de entender”
- “Puedo invertir mucho tiempo en programar al principio y luego es simplemente correr el robot”
- “Se puede verificar en el caso de mayor participación, pero no hay elementos suficientes en los otros casos”
- “Ni sí ni no... Habrá que ver cómo está organizado el curso y la planificación de actividades”

## A. CUESTIONARIO DE INDICADORES

---

- “Los gráficos indican un resultado a posteriori de la participación. Ésta puede haberse visto influida por múltiples factores (otros trabajos o entregas, circunstancias personales, etc.). No ofrece ninguna información fiable sobre el grado de planificación y gestión del tiempo del alumno”
- “Imposible si no se conoce la planificación previa de cada estudiante”
- “Sin saber en qué consiste el trabajo no se podría saber. Entiendo que no debería haber una única planificación/gestión que actúe como molde evaluador”
- “Al igual que dije anteriormente, por poner el foco en el resultado y no en el proceso. No puedo saber si ha habido una planificación (previa o durante la construcción del wiki). Sería interesante conocer el acuerdo al que ha llegado el equipo (o los acuerdos que se van tomando sobre la marcha) de cara a planificar su trabajo y gestión del tiempo, comprobando si efectivamente el resultado se ajusta a lo planificado”

### Liderazgo

Justificaciones para el sí:

- “Porque un estudiante destaca por encima de los otros dos, pudiendo deberse a un rol de liderazgo”
- “Puede distinguirse quien ha trabajado más (normalmente el líder, siempre y cuando se mida realmente la cantidad de trabajo en los gráficos, como he dicho anteriormente)”
- “Entiendo que sí sería posible, ya que el recorrido de cada gráfica nos indicaría si alguno de los participantes suele adelantarse a sus compañeros en las aportaciones, lo que podría significar un liderazgo en el proyecto.”
- “Mediante el nivel de participación en el propio tráfico de datos ...”
- “De acuerdo a la mayor o menor participación y trabajo = liderazgo”
- “Por la mayor participación ”

- “Puede observarse quién se ha implicado más en el proyecto y quiénes no lo han hecho”
- “Supongo que el que lleva el papel de líder reflejará un mayor tiempo de dedicación, tiempo que emplearía en revisar y comprobar las contribuciones de sus compañeros”
- “Al parecer los usuarios 2 y 3 aumentan su actividad conforme que vaya aumentando su actividad el usuario 1, que parece ser el más activo.”
- “Generalmente el alumno que adopta el papel de líder asume más responsabilidad y, por tanto, acaba aportando más en los trabajos por equipo.”
- “Lo que se ve es que el usuario 1 ha cargado casi todo el trabajo, pero insisto en que depende de qué trabajo es el que ha hecho. Puede haber sido buscar información y subirla, y el 3 analizarla que supone en definitiva más trabajo intelectual. Pero si hubiera tenido capacidad de liderazgo el trabajo hubiera sido más fluido”
- “Entiendo que sí. Supongo que sería aquel que más aportaciones haga de forma más continuada”
- “Se podría comprobar a partir del cual quien ha participado con mayores contribuciones y más valiosas, así como el seguimiento de sus ideas a través de las relaciones establecidas”

Justificaciones para el no:

- “No se puede establecer una relación directa entre la participación individual y la interacción con otros”
- “No veo forma de evaluar factores como el liderazgo a partir del número de ediciones, sin factores más cualitativos”
- “Los gráficos muestra qué estudiante ha sido el más activo, es decir el más trabajador. Ahora bien, de ahí deducir su liderazgo, no es plausible”

## A. CUESTIONARIO DE INDICADORES

---

- “Porque esta competencia no se refleja en los bytes aportados ni en su distribución en el tiempo, no puede conocerse quién promueve qué. Incluso los usuarios que menos información han aportado podrían haber centrado sus esfuerzos en ayudar a todos los demás, así como los que más información han aportado podrían haber actuado también como líderes predicando con el ejemplo”
- “Puedo invertir mucho tiempo en programar al principio y luego es simplemente correr el robot”
- “El liderazgo se puede ejercer de muchas maneras, pero rara vez viene dado por el peso directo del trabajo propio, que es de lo que dan idea los gráficos”
- “No creo que sería posible medir el liderazgo en esos gráficos, ya que no sabemos las motivaciones para participar en el trabajo (si los ha animado cualquier compañero y cuál de ellos)”
- “No veo cómo podría evaluarse dicha competencia a la vista de las gráficas, a menos que identifiquemos cantidad de participaciones con características de liderazgo, lo cual me parece un tanto arriesgado”
- “Porque no hay forma de saber quién dirige los trabajos. Aunque se podría inferir dada la mayor intervención de algún integrante, las razones para una mayor intervención pueden ser justo por una falta de liderazgo”
- “Nada me da datos al respecto”
- “Se muestra simplemente quién ha aportado más (cuantitativamente) y de forma más sostenida en el tiempo, pero no aporta información relativa a quién ha dirigido la discusión o ha mostrado más capacidad de liderazgo. Solo mide actividad y número de aportaciones, que puede ser parte de la evaluación de liderazgo pero no la define por completo”
- “No hay datos”
- “Porque no sabemos cómo ha sido el trabajo en equipo. Se necesita de observación”



- “El liderazgo no implica que un miembro del equipo haga más cosas visibles, podría ser lo contrario”
- “Necesitaría información sobre las interacciones entre los estudiantes”
- “Creo que esta competencia se observaría de forma directa en el trabajo en equipo. Sería la forma de ver quién ejerce el rol de líder, de organizar el trabajo, mediar ante las distintas opiniones y facilitar que el grupo llegue a un acuerdo.”
- “Porque no conocemos siquiera si ha existido comunicación entre ellos”
- “Inciendo nuevamente en las anteriores ideas plasmadas. ¿Cómo podría, por ejemplo, saber si un determinado miembro del equipo tiene capacidad para motivar a los demás para la consecución del objetivo común?. ¿Cómo podría saber quién toma las decisiones más importantes?. ¿Podríamos saber quién deja de ser un líder y cuándo otro asume dicho rol?”
- “A simple vista se aprecia quien ha hecho más aportaciones pero no quiere decir que haya sido el que ha liderado el trabajo. Valorar por suposiciones puede llevar a equívocos”
- “Creo que es difícil obtener información sobre quién ha liderado el grupo”
- “Los gráficos muestran el trabajo individual de los alumnos”
- “No es posible conocer las relaciones que se establecen entre ellos”



APÉNDICE

# B

## **Cuestionario complementario de la Actuación Avalada**

Este apéndice contiene el cuestionario complementario que se envió a los docentes que participaron en la actuación avalada. Las secciones de este apéndice son:

- Cuestionario (ver sección B.1)
- Resultados (ver sección B.2)

## **B. CUESTIONARIO COMPLEMENTARIO DE LA ACTUACIÓN AVALADA**

---

### **B.1 Cuestionario**

Los docentes que participaron en la actuación avalada recibieron un cuestionario complementario en la que marcaban para qué competencias genéricas considerarían los indicadores proporcionados por EvalCourse.

#### **A. Indicadores de accesos al campus virtual**

¿Qué competencia genérica podría evaluar con los indicadores de acceso al campus virtual?

- Trabajo en equipo
- Planificación y gestión del tiempo
- Razonamiento crítico y autocrítico
- Capacidad para identificar, plantear y resolver problemas
- Habilidades de interacción interpersonal
- Trabajo autónomo
- Otras

Explique o matice su respuesta si lo necesita y/o indique en qué otra competencia podría serle útil:

---

## B. Indicadores de los foros

¿Qué competencia genérica podría evaluar con los indicadores de los foros?

- Trabajo en equipo
- Planificación y gestión del tiempo
- Razonamiento crítico y autocrítico
- Capacidad para identificar, plantear y resolver problemas
- Habilidades de interacción interpersonal
- Trabajo autónomo
- Otras

Explique o matice su respuesta si lo necesita y/o indique en qué otra competencia podría serle útil:

---

## C. Indicadores del wiki

¿Qué competencia genérica podría evaluar con los indicadores del wiki?

- Trabajo en equipo
- Planificación y gestión del tiempo
- Razonamiento crítico y autocrítico
- Capacidad para identificar, plantear y resolver problemas
- Habilidades de interacción interpersonal
- Trabajo autónomo
- Otras

Explique o matice su respuesta si lo necesita y/o indique en qué otra competencia podría serle útil:

---

## D. Indicadores de las actividades

## **B. CUESTIONARIO COMPLEMENTARIO DE LA ACTUACIÓN AVALADA**

---

¿Qué competencia genérica podría evaluar con los indicadores de las actividades?

- Trabajo en equipo
- Planificación y gestión del tiempo
- Razonamiento crítico y autocrítico
- Capacidad para identificar, plantear y resolver problemas
- Habilidades de interacción interpersonal
- Trabajo autónomo
- Otras

Explique o matice su respuesta si lo necesita y/o indique en qué otra competencia podría serle útil:

---

## **E. Indicadores de los talleres**

¿Qué competencia genérica podría evaluar con los indicadores de los talleres?

- Trabajo en equipo
- Planificación y gestión del tiempo
- Razonamiento crítico y autocrítico
- Capacidad para identificar, plantear y resolver problemas
- Habilidades de interacción interpersonal
- Trabajo autónomo
- Otras

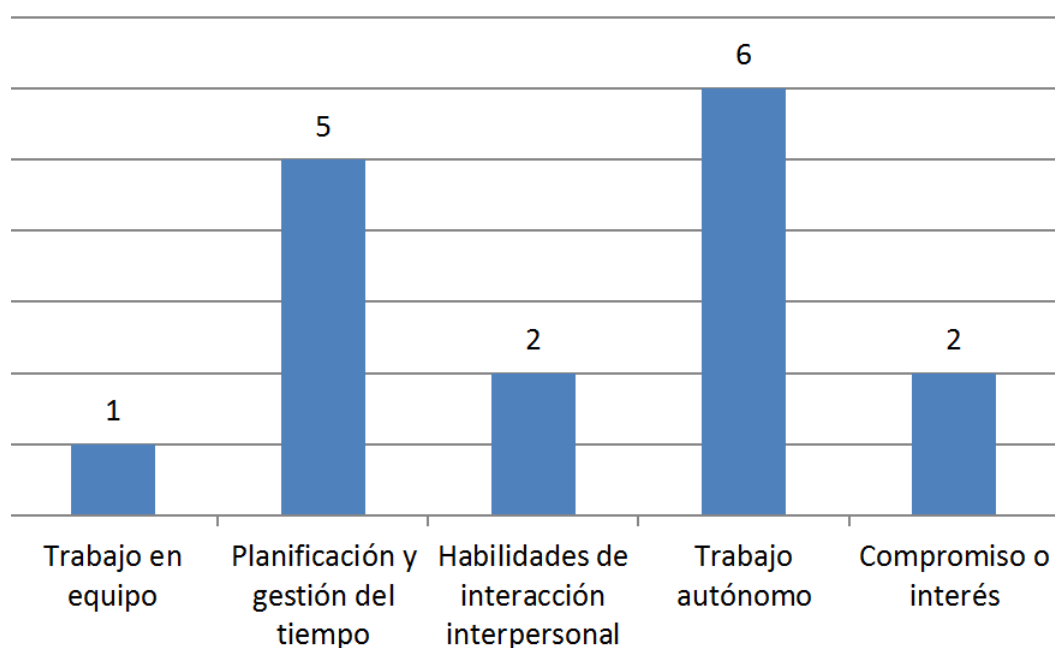
Explique o matice su respuesta si lo necesita y/o indique en qué otra competencia podría serle útil:

---

## B.2 Resultados

### A. Indicadores de accesos al campus virtual

El número de docentes que consideraría utilizar los indicadores de acceso al campus virtual para cada competencias genérica puede verse en la figura B.1.



**Figura B.1:** Docentes que considerarían los indicadores de accesos para evaluar cada competencia genérica de una población de 7

### Comentarios

- Interés por la asignatura. El interés y participación de los alumnos cambia drásticamente de un grado a otro e incluso de una facultad a otra. En la asignatura analizada del grado de Publicidad y Relaciones Públicas, impartido en la Facultad de Ciencias Sociales y de la Comunicación del Campus de Jerez parece que no son muy dados a visualizar y/o participar en el campus virtual de una asignatura, de hecho hay que fomentar su participación haciendo que ésta forme parte de la evaluación de la misma (mínimo un 10%). Con

## **B. CUESTIONARIO COMPLEMENTARIO DE LA ACTUACIÓN AVALADA**

---

este indicador, conociendo el número de días que se imparte la asignatura y obteniendo el número de veces que acceden al campus virtual, se puede determinar quiénes son alumnos muy activos, bastante activos, activos, poco activos y nada activos en la asignatura. Pudiéndose así evaluar el interés que puedan tener por la misma.

- Un mayor número de accesos al campus podría indicar quizás, grado de compromiso o interés.
- Es interesante observar la idoneidad del siguiente planteamiento: las capacidades de planificación y gestión de tiempo siguen una curva gaussiana desplazada hacia un menor número de accesos al campus, teniendo en cuenta factores como la participación práctica. Es decir, que una persona que accede muchas veces no necesariamente se planifica bien; diría que es al contrario.
- Planificación y gestión del tiempo, siempre y cuando se proporcionase el número de accesos por día, semana, etc.

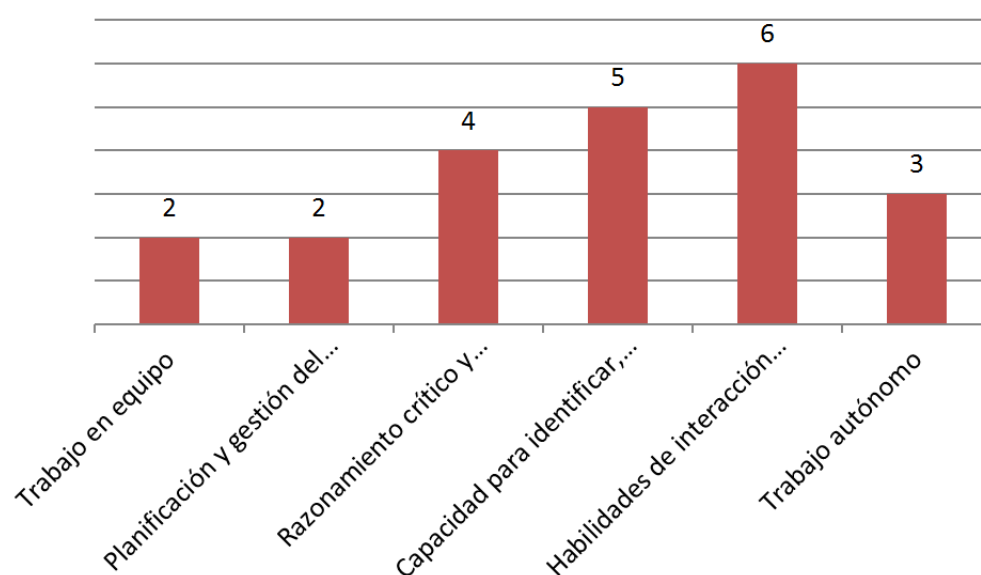
### **B. Indicadores de los foros**

El número de docentes que consideraría utilizar los indicadores de acceso al campus virtual para cada competencias genérica puede verse en la figura B.2.

#### **Comentarios**

- Uno de los foros de la asignatura se emplea para que los alumnos publiquen novedades, noticias, tutoriales, etc que sean de interés para la asignatura. Los alumnos identifican “elementos” que estén relacionados con la materia: bien sean noticias como tutoriales que puedan resolver algún problema práctico de la asignatura. Del mismo modo se plantean cuestiones al resto de compañeros fomentando la interacción interpersonal.
- La capacidad para identificar, plantear y resolver problemas sería posible marcarla si se analizara el contenido de los foros. Podemos suponer que una persona que responde puede resolver problemas, pero la identificación y el planteamiento no queda claro.





**Figura B.2:** Docentes que considerarían los indicadores de los foros para evaluar cada competencia genérica de un máximo

## C. Indicadores del wiki

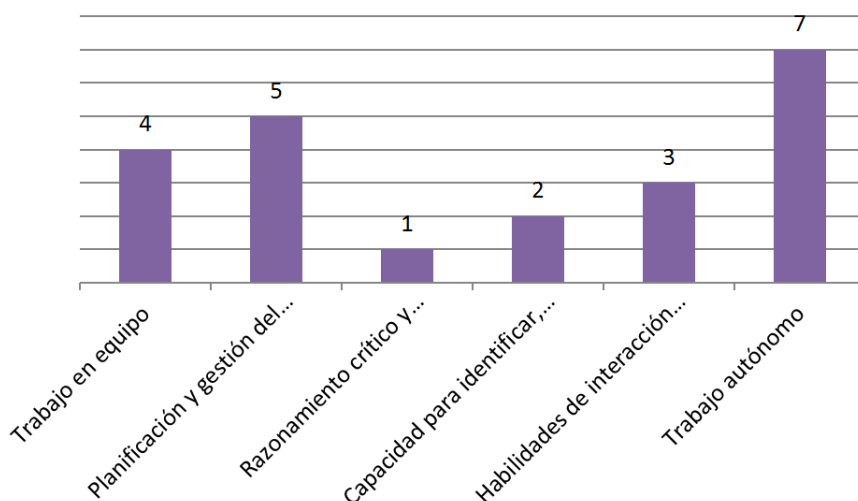
El número de docentes que consideraría utilizar los indicadores de acceso al campus virtual para cada competencias genérica puede verse en la figura B.3.

### Comentarios

- El wiki de los alumnos es un wiki individual para el seguimiento de una de las actividades a entregar. Cada vez que terminaban un paso del proyecto, iban completando una parte del wiki. Tenían todo el semestre para ir completando la actividad y wiki, lo que permite evaluar perfectamente cómo se ha planificado cada alumno en el curso, la capacidad para resolver los problemas y el trabajo autónomo del alumno.
- No me quedan claro lo que representan los ejes de las gráficas. Entiendo que se refiere a la cantidad de contribuciones (ordenadas) por semanas (abscisas). Refiero lo mismo en relación a la gestión del tiempo: no todo el que colabora mucho se gestiona de forma práctica. De hecho, un alumno “práctico” debería saber optimizar su tiempo, y generalmente la participación en una Wiki

## B. CUESTIONARIO COMPLEMENTARIO DE LA ACTUACIÓN AVALADA

---



**Figura B.3:** Docentes que considerarían los indicadores de los wikis para evaluar cada competencia genérica de un máximo de 7

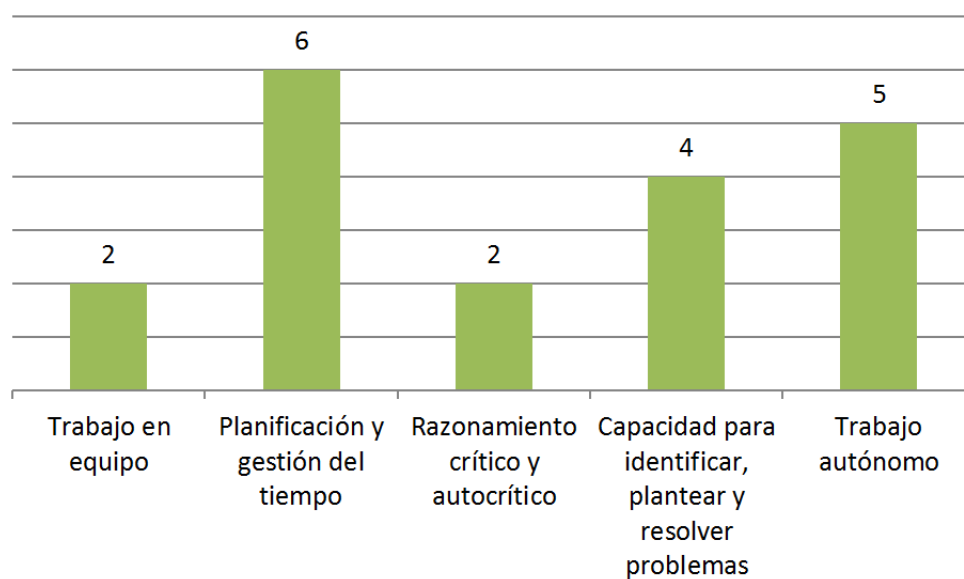
es superflua (salvo límites obligatorios). Las inquietudes personales y la auto-exigencia está muy relacionados con el rendimiento, pero no necesariamente con las habilidades. Esto es extensible al resto de los datos.

## D. Indicadores de las actividades

El número de docentes que consideraría utilizar los indicadores de acceso al campus virtual para cada competencias genérica puede verse en la figura B.4.

### Comentarios

- Las actividades a entregar en la asignatura eran de carácter individual, luego el hecho de la entrega de las mismas a tiempo permite evaluar perfectamente la planificación de los alumnos que la entregan (algunas actividades tenían una entrega optativa). En una de las actividades (optativas), en la que no se le facilitaban todos los pasos, permite evaluar que los alumnos han intentado identificar y resolver los problemas planteados y por supuesto, los que han participado, han fomentado su trabajo autónomo en la asignatura.
- Si se proponen actividades en grupo, éstas podrían medir el trabajo en equipo.



**Figura B.4:** Docentes que considerarían los indicadores de las actividades para evaluar cada competencia genérica de un máximo de 7

- Planificación y gestión del tiempo, se presupone que es posible acabar las tareas en un tiempo razonable sin “quemar” al alumno.
- Trabajo en equipo, siempre y cuando se considere en los resultados las tareas con la opción de Moodle de “entrega por grupo”.

### E. Indicadores de los talleres

Sólo hubo una profesora que utilizó los talleres y en concreto los consideró para evaluar las competencias de la capacidad para identificar, plantear y resolver problemas, las habilidades de interacción interpersonal y el trabajo autónomo.



# **Evaluación del método de evaluación (DBA) y del lenguaje para diseñar evaluaciones (SASQL)**

Este apéndice contiene el cuestionario utilizado para la evaluación del método DBA y del lenguaje SASQL. Además, en un segundo apartado se analizan las respuestas recogidas. Las secciones de este apéndice son:

- Cuestionario (ver sección C.1)
- Resultados (ver sección C.2)

## **C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)**

---

### **C.1 Cuestionario**

#### **A. Sexo**

- Hombre
- Mujer

#### **B. Rama**

- Arte y humanidades
- Ciencias sociales y jurídicas
- Ciencias
- Ciencias de la salud
- Ingeniería y arquitectura

### C. Cuestiones sobre el método DBA

En este método, los docentes diseñan una evaluación de los estudiantes a partir de la información contenida en el registro del entorno virtual de aprendizaje y evalúan los resultados. El docente podrá aceptar y utilizar los resultados, descartarlos o redefinir una nueva evaluación para contrastar los resultados anteriores.

#### C.1 ¿Considera el método DBA adecuado para obtener automáticamente indicadores de los entornos virtuales de aprendizaje?

- Sí
- No
- Ns/nc

#### C.2 ¿Cuál de las siguientes características considera que aporta el método DBA?

Marque una opción para cada característica siendo (1) totalmente de acuerdo, (2) de acuerdo, (3) ni acuerdo ni en desacuerdo, (4) en desacuerdo y (5) totalmente en desacuerdo.

CARACTERISTICAS	(1)	(2)	(3)	(4)	(5)
Objetividad					
Adaptabilidad					
Sistematicidad					
Flexibilidad					
Fiabilidad					

**Tabla C.1:** Selección de características del método DBA

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

### D. Cuestiones sobre el lenguaje SASQL

Con este lenguaje los docentes pueden diseñar evaluaciones a partir de la información contenida en el registro de aprendizaje virtual.

**D.1 ¿Considera que el lenguaje presentado es útil para diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos virtuales de aprendizaje?**

- Sí
- No
- Ns/Nc

**D.2 ¿Cuáles de las siguientes características considera que aporta el lenguaje presentado?**

Marque una opción para cada característica siendo (1) totalmente de acuerdo, (2) de acuerdo, (3) ni acuerdo ni en desacuerdo, (4) en desacuerdo y (5) totalmente en desacuerdo.

CARACTERISTICAS	(1)	(2)	(3)	(4)	(5)
Facilidad de aprendizaje					
Eficacia					
Ahorra tiempo					
Escalabilidad					
Capacidad de reutilización					
Fiabilidad					

**Tabla C.2:** Selección de características del lenguaje SASQL



D.3 Tomando como ejemplo la siguiente consulta del DSL, utilizada para obtener información de la interacción de los estudiantes en el foro del campus virtual:

```
Evidence interaccion_foro:
  get students
  show interaction
  in forum.
```

- ¿Considera entendible la consultas del DSL?
- Sí
  - No
  - Ns/Nc

D.4 Tomando como ejemplo los resultados que se obtienen de la consulta anterior:

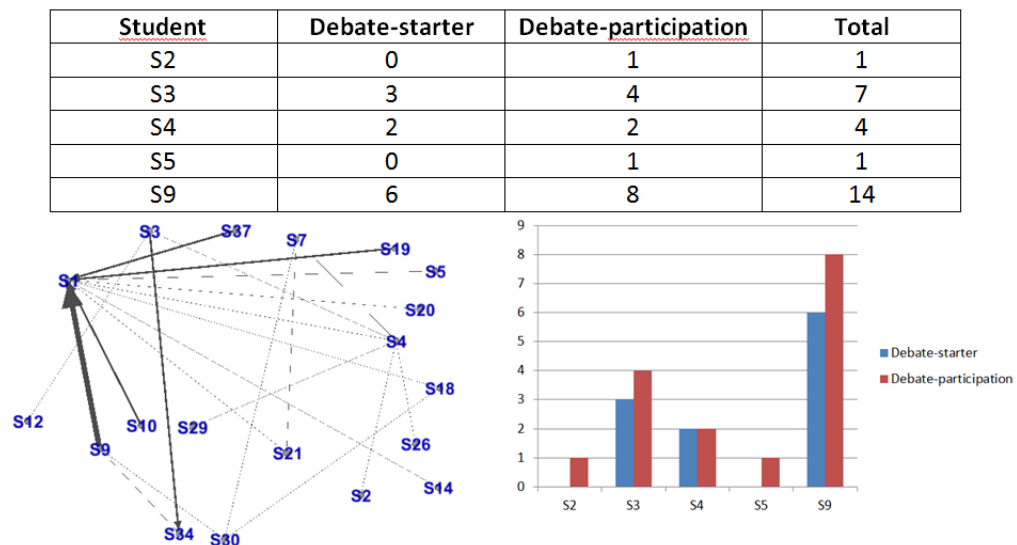


Figura C.1: Resultados consulta

### **C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)**

---

¿Considera los resultados entendibles?

- Sí
- No
- Ns/Nc

## C.2 Resultados

En esta sección se muestran y analizan en detalle los resultados del cuestionario.

### C.2.1 Participantes

En este cuestionario participaron docentes que asistieron a las Jornadas de Innovación Docente 2016 de la Universidad de Cádiz. Se realizó una presentación del método y la herramienta, y tras un turno de preguntas, se pasó el cuestionario mostrado en el apartado anterior.

En total fueron completados 31 cuestionarios, distribuidos tal y cómo se resume en la tabla C.3.

Rama Sexo	Hombre	Mujer	Ns/Nc	Total
Arte y humanidades	1	4	0	5
Ciencias	3	2	0	5
Ciencias de la salud	6	0	0	6
Ciencias sociales y jurídicas	2	2	1	5
Ingeniería y arquitectura	6	4	0	10

**Tabla C.3:** Resumen de distribución de participantes por rama y sexo

Para afinar en la independencia de las respuestas dadas por los docentes se agruparon sus ramas con respecto a las ramas principales del conocimiento:

- *EHSE, Earth & Health Sciences and Engineering*: ciencias, ciencias de la salud e ingeniería y arquitectura.
- *SSH, Social Sciences and Humanities*: arte y humanidades y ciencias sociales y jurídicas.

### C.2.2 Evaluación del método DBA

Con respecto a la pregunta sobre la idoneidad del método DBA, el resumen de las respuestas puede verse en la tabla C.4.

### C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

	Rama	Sí	No	Total
SSH	Arte y humanidades	4	1	5
	Ciencias sociales y jurídicas	3	2	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	9	1	10

**Tabla C.4:** Consideración de idoneidad del método DBA por ramas

Un 84% de los encuestados (26 de 31) consideraron el método DBA adecuado. Para dar validez a los resultados es necesario determinar la independencia de los mismos.

En primer lugar, para determinar si existe relación entre la respuesta de cada docente y su rama se define la siguiente hipótesis nula:

$H_0$ : La consideración por parte de un docente de que el método DBA sea o no válido es independiente de la rama de dicho docente

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,6142**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.

En segundo lugar, para determinar si existe relación entre la respuesta de cada docente y su rama principal se define la siguiente hipótesis nula:

$H_0$ : La consideración por parte de un docente de que el método DBA sea o no válido es independiente de si dicho docente es de EHSE o de SSH

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,2955**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.

#### Características del método DBA

A continuación se preguntó a los encuestados sobre su consideración acerca de en qué medida el método DBA satisface cada una de las características que se espe-

## C.2 Resultados

ran del mismo. Estas características son *objetividad*, *adaptabilidad*, *sistematicidad*, *flexibilidad* y *fiabilidad*. Para su evaluación se preguntó sobre ellos en una escala Likert de cinco puntos: totalmente de acuerdo, de acuerdo, ni de acuerdo ni en desacuerdo, en desacuerdo y totalmente en desacuerdo.

En la tabla C.5 puede verse el resumen de las respuestas dadas.

Características	Totalmente de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Totalmente en desacuerdo
Objetividad	23%	23%	32%	19%	3%
Adaptabilidad	16%	45%	23%	13%	3%
Sistematicidad	40%	30%	20%	10%	0%
Flexibilidad	19%	42%	26%	13%	0%
Fiabilidad	6%	39%	26%	23%	6%

**Tabla C.5:** Evaluación de las características del método DBA

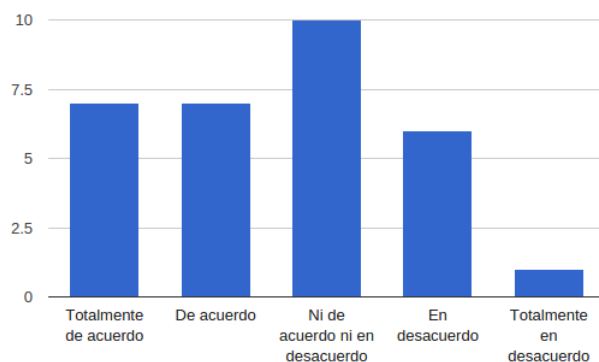
En general los docentes muestran su conformidad para las características. En primer lugar, para la adaptabilidad, la sistematicidad y la flexibilidad, el nivel de aceptación de los docentes para el método está entre el 61 y el 70%. En segundo lugar, para las características de la fiabilidad y la objetividad la suma de las dos primeras opciones está en torno al 45%.

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

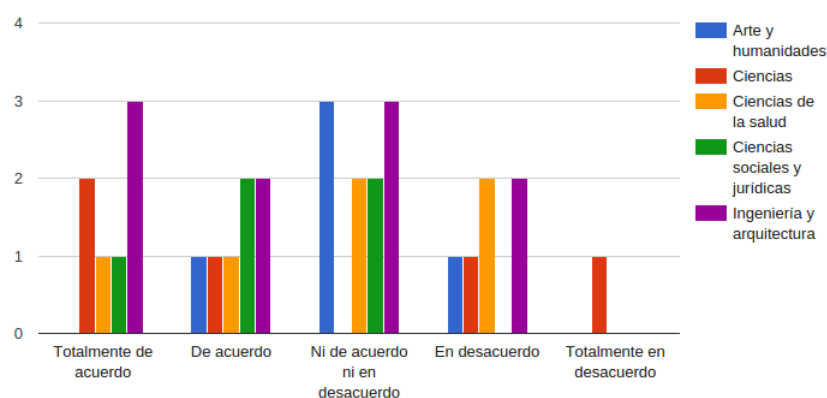
### Objetividad

El resumen de respuestas para la característica de la objetividad puede verse en las figuras C.2 y C.3.



**Figura C.2:** Resumen global de respuestas para la característica de la objetividad

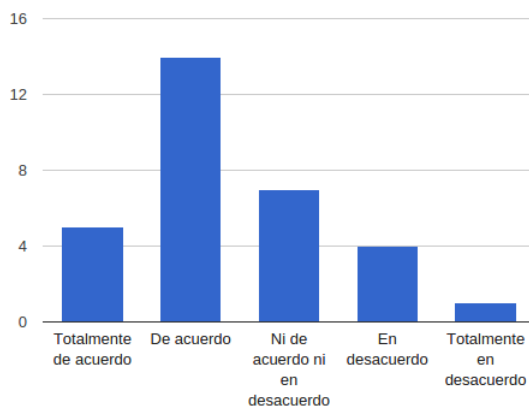
Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el método es objetivo (14 de 31). Por otro lado, 10 docentes mostraron una posición neutral con respecto a esta característica, mientras que 7 estuvieron en desacuerdo. Los datos no son significativos en lo que respecta a la respuesta respecto de la rama de cada docente, ya que como puede observarse en la figura C.3, no hay una tendencia definida.



**Figura C.3:** Resumen global de respuestas para la característica de la objetividad por rama

### Adaptabilidad

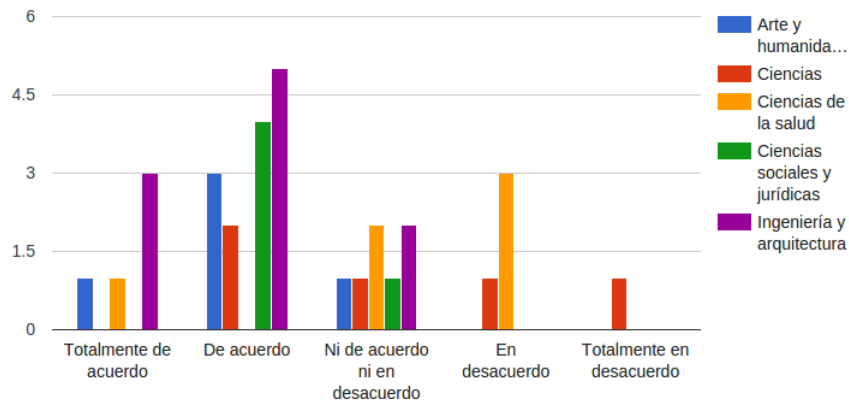
El resumen de respuestas para la característica de la adaptabilidad puede verse en las figuras C.4 y C.5.



**Figura C.4:** Resumen global de respuestas para la característica de la adaptabilidad

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el método proporciona adaptabilidad (19 de 31). Por otro lado, 7 docentes mostraron una posición neutral con respecto a esta característica, mientras que 5 estuvieron en desacuerdo. Cabe destacar que los docentes que estuvieron en desacuerdo fueron docentes de ciencias o ciencias de la salud.

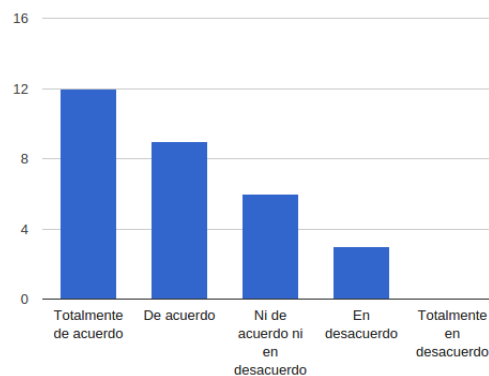
## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)



**Figura C.5:** Resumen global de respuestas para la característica de la adaptabilidad por rama

### Sistematicidad

El resumen de respuestas para la característica de la sistematicidad puede verse en las figuras C.6 y C.7.

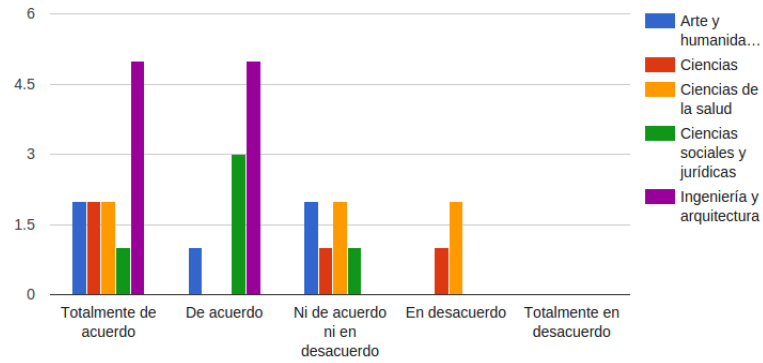


**Figura C.6:** Resumen global de respuestas para la característica de la sistematicidad

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el método proporciona sistematicidad (21 de 31). Por otro lado, 6 docentes mostraron una posición neutral con respecto a esta característica, mientras que 3 estuvieron en desacuerdo. Puede destacarse de nuevo que los docentes que estuvieron en desacuerdo fueron docentes de ciencias o ciencias de la



salud, mientras que todos los docentes de Ingeniería y arquitectura mostraron su conformidad.



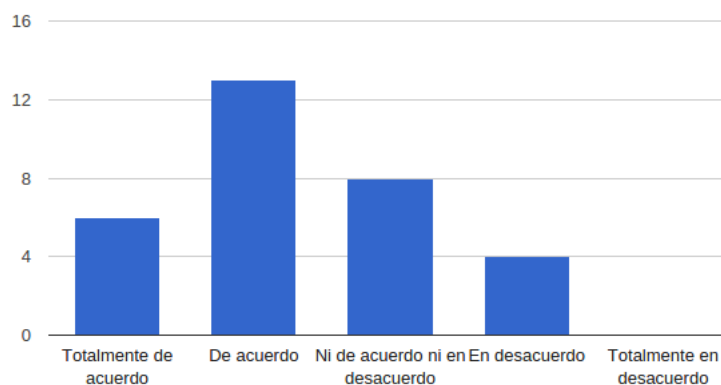
**Figura C.7:** Resumen global de respuestas para la característica de la sistematicidad por rama

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

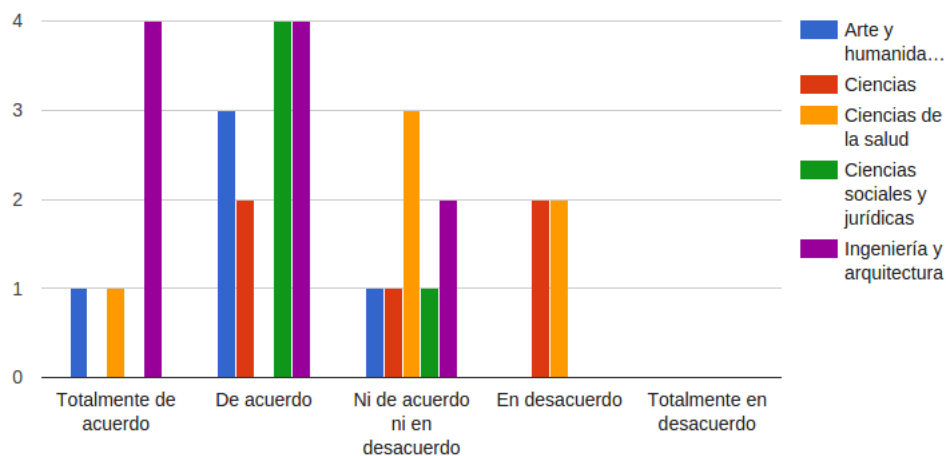
### Flexibilidad

El resumen de respuestas para la característica de la flexibilidad puede verse en las figuras C.8 y C.9.



**Figura C.8:** Resumen global de respuestas para la característica de la flexibilidad

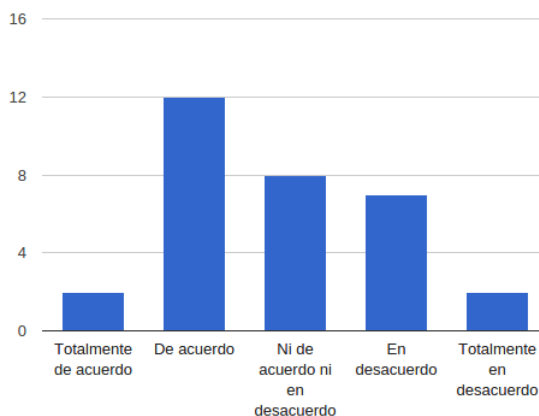
Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el método proporciona flexibilidad (19 de 31). Además, 8 docentes mostraron una posición neutral con respecto a esta característica, mientras que 4 estuvieron en desacuerdo. También los docentes que estuvieron en desacuerdo fueron docentes de ciencias o ciencias de la salud.



**Figura C.9:** Resumen global de respuestas para la característica de la flexibilidad por rama

### Fiabilidad

El resumen de respuestas para la característica de la fiabilidad puede verse en las figuras C.10 y C.11.

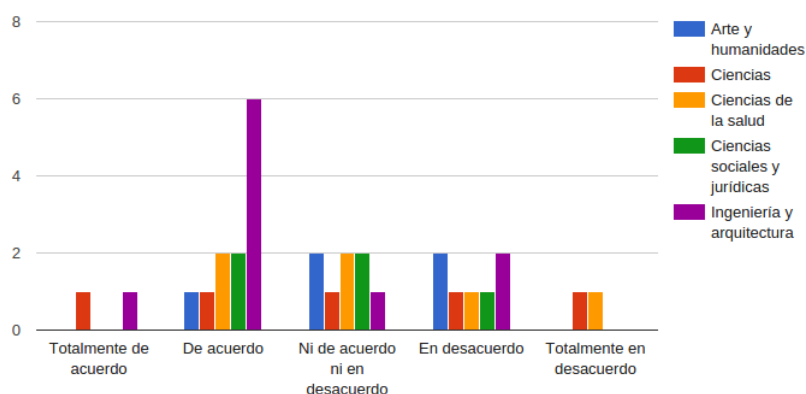


**Figura C.10:** Resumen global de respuestas para la característica de la fiabilidad

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el método es fiable (14 de 31). Además, 8 docentes mostraron una posición neutral con respecto a esta característica, mientras que 9 estuvieron

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

en desacuerdo. Los datos no son significativos en lo que respecta a la respuesta respecto de la rama de cada docente, ya que como puede observarse en la figura C.11, no hay una tendencia definida.



**Figura C.11:** Resumen global de respuestas para la característica de la fiabilidad por rama

### C.2.3 Evaluación del DSL

Con respecto a la pregunta sobre la idoneidad del DSL, el resumen de las respuestas puede verse en la tabla C.6.

	Rama	Sí	No	Total
SSH	Arte y humanidades	4	1	5
	Ciencias sociales y jurídicas	3	2	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	9	1	10

**Tabla C.6:** Consideración de idoneidad del DSL por ramas

Un 84% de los encuestados (26 de 31) consideraron el DSL adecuado. Para dar validez a los resultados es necesario determinar la independencia de los mismos.

En primer lugar, para determinar si existe relación entre la respuesta de cada docente y su rama se define la siguiente hipótesis nula:

*H<sub>0</sub>: La consideración por parte de un docente de que el DSL sea o no válido es independiente de la rama de dicho docente*

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,6142**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.

En segundo lugar, para determinar si existe relación entre la respuesta de cada docente y su rama principal se define la siguiente hipótesis nula:

*H<sub>0</sub>: La consideración por parte de un docente de que el DSL sea o no válido es independiente de si dicho docente es de EHSE o de SSH*

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,2955**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.

### Características del DSL

A continuación se preguntó a los encuestados sobre su consideración acerca de en qué medida el DSL satisface cada una de las características que se esperan del mismo. Estas características son *facilidad de aprendizaje, eficacia, ahorra tiempo, escalabilidad, capacidad de reutilización y fiabilidad*. Para su evaluación se preguntó sobre ellos en una escala Likert de cinco puntos: totalmente de acuerdo, de acuerdo, ni de acuerdo ni en desacuerdo, en desacuerdo y totalmente en desacuerdo.

En la tabla C.7 puede verse el resumen de las respuestas dadas.

### C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

Características	Totalmente de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Totalmente en desacuerdo
Facilidad de aprendizaje	19%	32%	39%	7%	3%
Eficacia	13%	45%	26%	16%	0%
Ahorra tiempo	32%	23%	32%	13%	0%
Escalabilidad	23%	23%	35%	19%	0%
Capacidad de reutilización	26%	32%	29%	6,5%	6,5%
Fiabilidad	6%	29%	42%	23%	0%

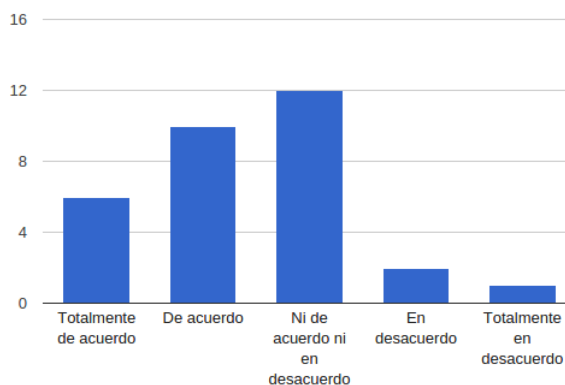
**Tabla C.7:** Evaluación de las características del DSL

Puede decirse que el rechazo al hecho de que el DSL proporcione dichas características es bajo, pues para todas las características es inferior al 20% (excepto para la fiabilidad que es el 23%).

Sin embargo, es un hecho que el nivel de aceptación (sumando los de acuerdo con los totalmente de acuerdo) es levemente superior al 50%, habiendo también un grupo nutrido de docentes que se sitúan en el punto neutro.

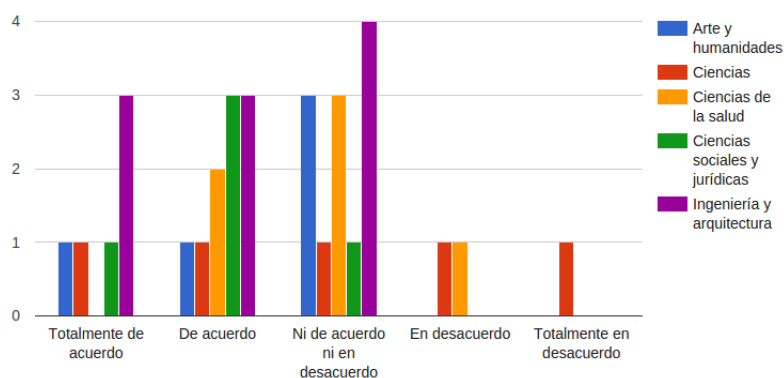
### Facilidad de aprendizaje

El resumen de respuestas para la característica de la facilidad de aprendizaje del DSL puede verse en las figuras C.12 y C.13.



**Figura C.12:** Resumen global de respuestas para la característica de la facilidad de aprendizaje

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el DSL es fácil de aprender (16 de 31). Además, 12 docentes mostraron una posición neutral con respecto a esta característica del DSL, mientras que 3 estuvieron en desacuerdo. Los docentes que estuvieron en desacuerdo fueron nuevamente docentes de ciencias o ciencias de la salud.



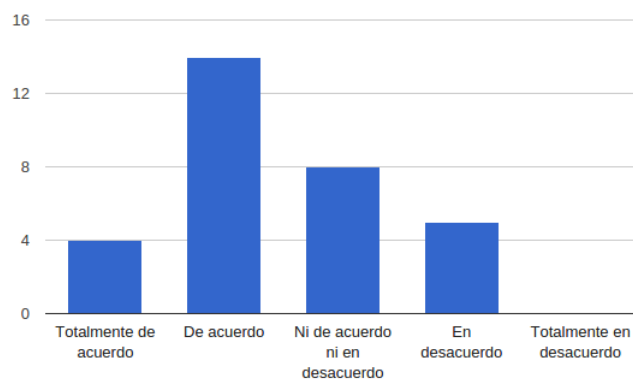
**Figura C.13:** Resumen global de respuestas para la característica de la facilidad de aprendizaje por rama

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

### Eficacia

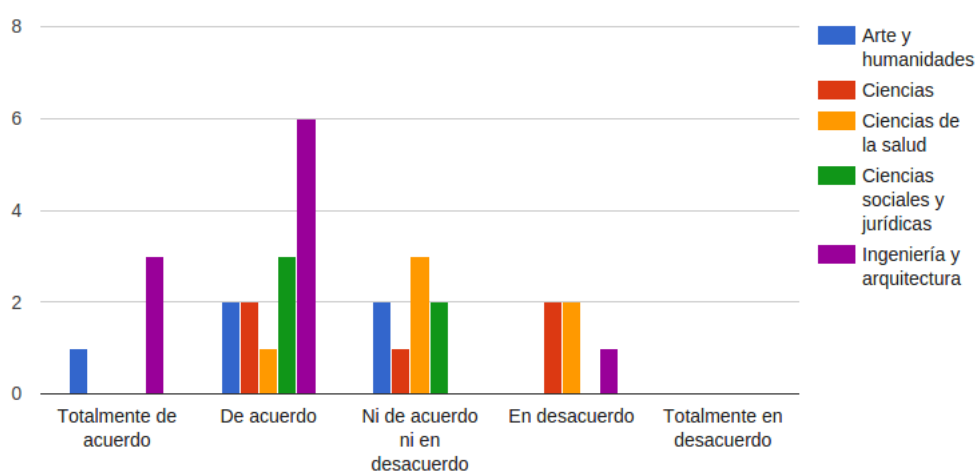
El resumen de respuestas para la característica de la eficacia del DSL puede verse en las figuras C.14 y C.15.



**Figura C.14:** Resumen global de respuestas para la característica de la eficacia

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el DSL es eficaz (18 de 31). Además, 8 docentes mostraron una posición neutral con respecto a esta característica, mientras que 5 estuvieron en desacuerdo. Nada que destacar con respecto a la distribución de las respuestas por rama que puede verse en la figura C.15.

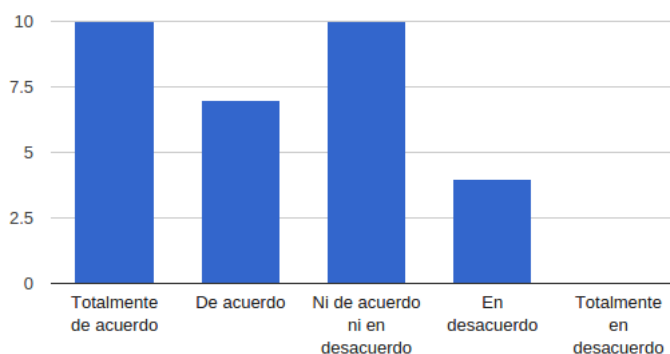




**Figura C.15:** Resumen global de respuestas para la característica de la eficacia por rama

### Ahorra tiempo

El resumen de respuestas para la característica del ahorro del tiempo del DSL puede verse en las figuras C.16 y C.17.



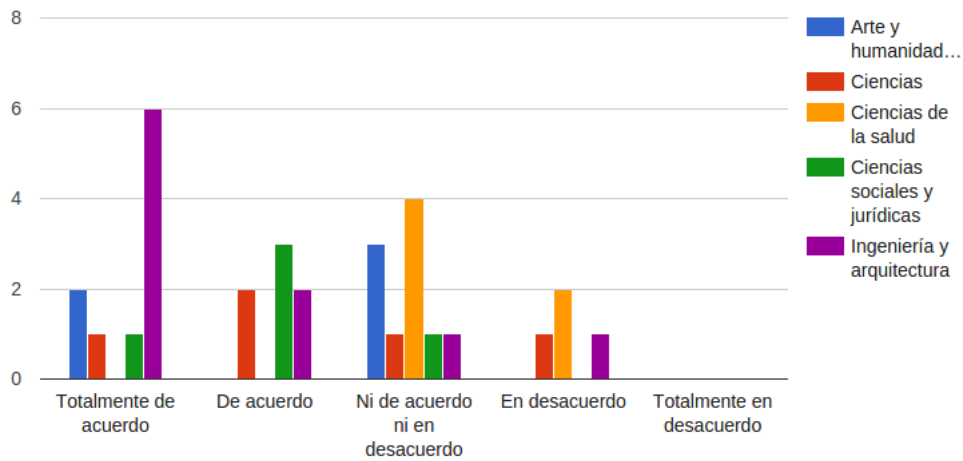
**Figura C.16:** Resumen global de respuestas para la característica del ahorro del tiempo

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que usando el DSL se ahorra tiempo (17 de 31). Además, 10 docentes mostraron una posición neutral con respecto a esta característica, mientras

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

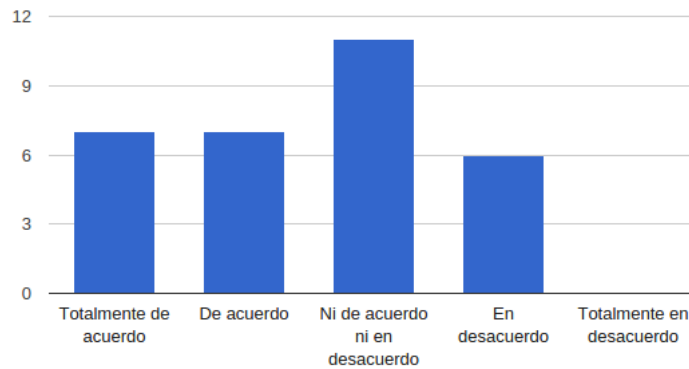
que 4 estuvieron en desacuerdo. Nada que destacar con respecto a la distribución de las respuestas por rama que puede verse en la figura C.17.



**Figura C.17:** Resumen global de respuestas para la característica del ahorro del tiempo

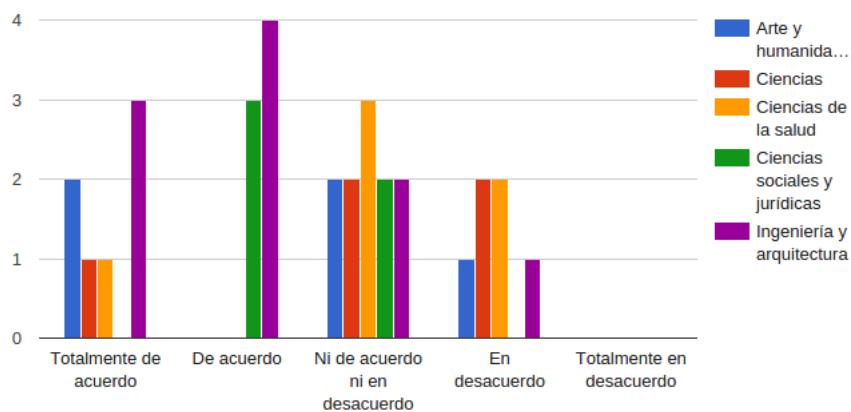
**Escalabilidad**

El resumen de respuestas para la característica de la escalabilidad del DSL puede verse en las figuras C.18 y C.19.



**Figura C.18:** Resumen global de respuestas para la característica de la escalabilidad

Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con el hecho de que el método es escalable (14 de 31). Además, 10 docentes mostraron una posición neutral con respecto a esta característica, mientras que 6 estuvieron en desacuerdo. Nada que destacar con respecto a la distribución de las respuestas por rama que puede verse en la figura C.19.



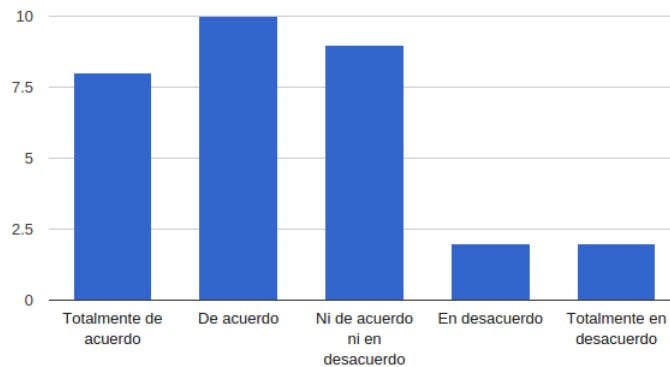
**Figura C.19:** Resumen global de respuestas para la característica de la escalabilidad por rama

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

---

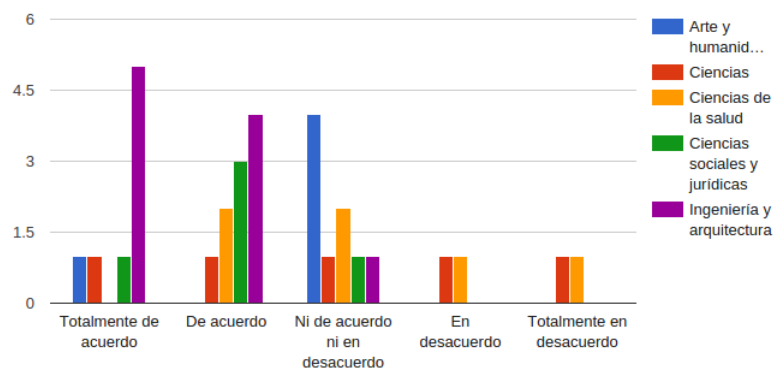
### Capacidad de reutilización

El resumen de respuestas para la característica de la capacidad de reutilización del DSL puede verse en las figuras C.20 y C.21.



**Figura C.20:** Resumen global de respuestas para la característica de la capacidad de reutilización

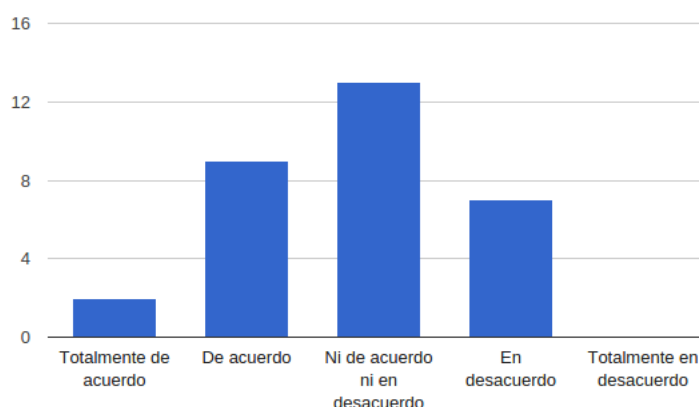
Hay una mayoría de docentes que están en mayor o menor medida de acuerdo con la capacidad de reutilización que proporciona el DSL (18 de 31). Además, 9 docentes mostraron una posición neutral con respecto a esta característica, mientras que 4 estuvieron en desacuerdo. En este caso, los únicos docentes que estuvieron en desacuerdo fueron nuevamente docentes de ciencias o ciencias de la salud.



**Figura C.21:** Resumen global de respuestas para la característica de la capacidad de reutilización por rama

### Fiabilidad

El resumen de respuestas para la característica de la fiabilidad del DSL puede verse en las figuras C.22 y C.23.



**Figura C.22:** Resumen global de respuestas para la característica de la fiabilidad

Hay una mayoría de docentes que están en una posición neutra (13 de 31) con respecto a la fiabilidad que consideran que proporciona el DSL. Además, 11 docentes mostraron estar en mayor o menor medida de acuerdo con respecto a esta característica, mientras que 7 estuvieron en desacuerdo.

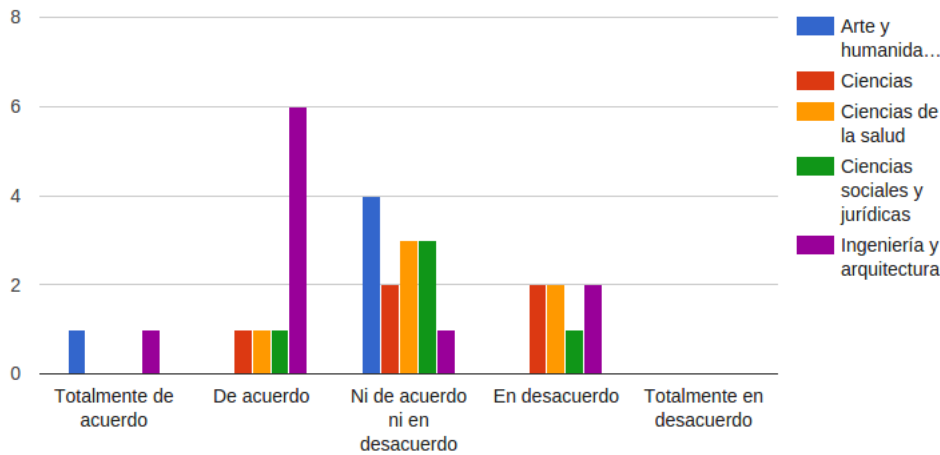
#### C.2.4 Evaluación de la sintaxis del DSL (consulta tipo)

Con respecto a la pregunta sobre la entendibilidad de la sintaxis del DSL, el resumen de las respuestas puede verse en la tabla C.4.

	Rama	Sí	No	Total
SSH	Arte y humanidades	2	3	5
	Ciencias sociales y jurídicas	3	2	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	9	1	10

**Tabla C.8:** Consideración de la consulta como entendible o no por ramas

## C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)



**Figura C.23:** Resumen global de respuestas para la característica de la fiabilidad por rama

Un 77% de los encuestados (24 de 31) consideraron la sintaxis DSL entendible. Para dar validez a los resultados es necesario determinar la independencia de los mismos.

En primer lugar, para determinar si existe relación entre la respuesta de cada docente y su rama se define la siguiente hipótesis nula:

$H_0$ : La consideración por parte de un docente de que la sintaxis del DSL sea o no entendible es independiente de la rama de dicho docente

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,121**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.

En segundo lugar, para determinar si existe relación entre la respuesta de cada docente y su rama principal se define la siguiente hipótesis nula:

$H_0$ : La consideración por parte de un docente de que la sintaxis del DSL sea o no entendible es independiente de si dicho docente es de EHSE o de SSH

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,021**, superior al umbral de significación del 0,005.

Por lo que se concluye que, con una significación del 5%, los datos son independientes.

### C.2.5 Evaluación de los resultados

Con respecto a la pregunta sobre la entendibilidad de los resultados devueltos por el DSL, el resumen de las respuestas puede verse en la tabla C.4.

	Rama	Sí	No	Total
SSH	Arte y humanidades	3	2	5
	Ciencias sociales y jurídicas	0	5	5
EHSE	Ciencias	5	0	5
	Ciencias de la salud	5	1	6
	Ingeniería y arquitectura	7	3	10

**Tabla C.9:** Consideración de los resultados como entendibles o no por ramas

Un 64% de los encuestados (20 de 31) consideraron los resultados entendibles. Para dar validez a los resultados del cuestionario es necesario determinar la independencia de los mismos.

En primer lugar, para determinar si existe relación entre la respuesta de cada docente y su rama se define la siguiente hipótesis nula:

$H_0$ : La consideración por parte de un docente de que los resultados sean o no entendibles es independiente de la rama de dicho docente

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,010**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.

En segundo lugar, para determinar si existe relación entre la respuesta de cada docente y su rama principal se define la siguiente hipótesis nula:

$H_0$ : La consideración por parte de un docente de que los resultados sean o no entendibles es independiente de si dicho docente es de EHSE o de SSH

### **C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)**

---

Para contrastar la hipótesis se utilizó el test de Fisher, no pudiéndose rechazar la hipótesis con un **p-valor de 0,013**, superior al umbral de significación del 0,005. Por lo que se concluye que, con una significación del 5%, los datos son independientes.



APÉNDICE

# D

## Listado de competencias genéricas

A continuación se muestra el listado de competencias genéricas (tabla D.1). La primera columna representa el identificador utilizado en el documento para cada competencia, la segunda columna representa la denominación de la competencia y la tercera columna la descripción.

## **D. LISTADO DE COMPETENCIAS GENÉRICAS**

---

Esta página se ha dejado en blanco intencionadamente.

ID	COMPETENCIA	DESCRIPCIÓN
CHA	Adaptación al cambio	Capacidad adaptarse y actuar ante nuevas situaciones [52].
ANA	Análisis	Capacidad de abstracción, análisis y síntesis [52].
LLL	Aprendizaje permanente	Aprendizaje formal, no formal e informal que tiene lugar a lo largo de la vida y con el fin de alcanzar el máximo desarrollo de la personalidad y de las destrezas profesionales [99].
QUA	Calidad	Capacidad para evaluar y mantener la calidad del trabajo producido [52].
COM	Comunicación	Habilidad para comunicarse de manera tanto oral como escrita en la lengua materna [52].
KNO	Conocimiento	Capacidad de aplicar el conocimiento en situaciones prácticas [52].
CRE	Creatividad	Capacidad para crear nuevas ideas [52].
CUL	Cultural	Aprecio y respeto por la diversidad y la multiculturalidad [52].
ENT	Emprendimiento	Capacidad para tomar la iniciativa y espíritu de empresa [52].
MAN	Gestión de proyectos	Habilidades para diseñar y gestionar proyectos [52].
INT	Habilidades interpersonales	Capacidad de la persona para comunicarse e interactuar con otras personas [52].
INV	Investigación	Capacidad para llevar a cabo la investigación en un nivel apropiado [52].
LEN	Lengua extranjera	Capacidad de los estudiantes para comunicar sus ideas en un segundo idioma [50].
LEA	Liderazgo	Habilidad para motivar a la gente y conducirlos hacia un objetivo común [52].
CRI	Pensamiento crítico	Habilidad para interpretar, analizar y evaluar ideas y argumentos [46].
PLA	Planificación y gestión del tiempo	Capacidad de planificar y gestionar el tiempo de manera efectiva [52].
PRO	Resolución de problemas	Habilidad para identificar, plantear y resolver problemas [52].
RSP	Responsabilidad	Capacidad para actuar con responsabilidad social y conciencia cívica [52].
TIC	TIC	Habilidades en el uso de las tecnologías de la información y de la comunicación [52].
DEC	Toma de decisiones	Capacidad para tomar decisiones razonadas [52].
AUT	Trabajo autónomo	Capacidad para trabajar de forma autónoma [52].
TMW	Trabajo en equipo	Trabajo realizado por un conjunto de personas en aras de un objetivo común [35].

**Tabla D.1:** Competencias genéricas



# Bibliografía

- [1] F. Achcaoucaou, L. Guitart-Tarrés, P. Miravittles-Matamoros, A. Núñez-Carballosa, M. Bernardo, and A. Bikfalvi. Competence assessment in higher education: A dynamic approach. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 24(4):454–467, 2014. 35, 37, 38, 46
- [2] A. Al-Ajlan and H. Zedan. Why moodle. In *Future Trends of Distributed Computing Systems, 2008. FTDCS'08. 12th IEEE International Workshop on*, pages 58–64. IEEE, 2008. 14
- [3] C. Alario-Hoyos, M. L. Bote-Lorenzo, E. Gómez-Sánchez, J. I. Asensio-Pérez, G. Vega-Gorgojo, and A. Ruiz-Calleja. Glue!: An architecture for the integration of external tools in virtual learning environments. *Computers & Education*, 60(1):122–137, 2013. 4
- [4] T. Anderson and J. Shattuck. Design-based research a decade of progress in education research? *Educational researcher*, 41(1):16–25, 2012. 6, 71, 72
- [5] M. André, M. G. Baldoquín, and S. T. Acuña. Formal model for assigning human resources to teams in software projects. *Information and Software Technology*, 53(3):259–275, 2011. 41, 42, 43, 47, 61
- [6] F. G. Andrés. Los beneficios de la tecnología en la educación. <http://www.labrechadigital.org/labrecha/Articulos/los-beneficios-de-la-tecnologia-en-la-educacion.html>, 2015. Accessed: 2016-01-11. 2

## BIBLIOGRAFÍA

---

- [7] E. Arno-Macia and C. Rueda-Ramos. Promoting reflection on science, technology, and society among engineering students through an eap online learning environment. *Journal of English for Academic Purposes*, 10(1):19–31, 2011. 35, 37, 38, 48
- [8] A. A. Aziz, A. Mohamed, N. Arshad, S. Zakaria, and M. S. Masodi. Appraisal of course learning outcomes using rasch measurement: a case study in information technology education. *International Journal of Systems Applications, Engineering & Development*, 4(1):164–172, 2007. 40, 46
- [9] A. Balderas, A. Berns, M. Palomo-Duarte, J. M. Dodero, R. Gómez-Sánchez, and I. Ruiz-Rube. A domain specific language to retrieve objective indicators for foreign language learning in virtual worlds. In *Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality*, pages 675–680. ACM, 2015. 108, 109, 138, 139, 146, 148, 149
- [10] A. Balderas, J. M. Dodero, M. Palomo-Duarte, and I. Ruiz-Rube. A domain specific language for online learning competence assessments. *International Journal of Engineering Education - Special issue on Innovative Methods of Teaching Engineering*, 31(3):851–862, 2015. 101, 109, 136, 139, 146, 149
- [11] A. Balderas, A. Galan-Pinero, J. A. Caballero Hernandez, G. Rodriguez Gomez, J. Dodero, and M. Palomo-Duarte. Domain-driven competence assessment in virtual learning environments. application to planning and time management skills. In *Computers in Education (SIIE), 2014 International Symposium on*, pages 121–126. IEEE, 2014. 109, 138, 147
- [12] A. Balderas, M. Palomo-Duarte, J. M. Dodero, and I. R. Rube. Qualitative assessment of wiki-based learning processes. In *Proceedings of SPDECE-2012. Ninth multidisciplinary symposium on the design and evaluation of digital content for education*, pages 161–172, 2012. 109, 137, 139, 147, 149
- [13] A. Balderas, I. Ruiz-Rube, J. M. Dodero, M. Palomo-Duarte, and A. Berns. A generative computer language to customize online learning assessments.

- Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8095 LNCS:591–592, 2013. 91, 137, 139, 147, 149
- [14] A. Balderas, I. Ruiz-Rube, M. Palomo-Duarte, and J. M. Doderó. A generative computer language to customize online learning assessments. In *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, pages 141–147. ACM, 2013. 109, 138, 139, 147, 149
- [15] M. Bedek, S. A. Petersen, and T. Heikura. From behavioral indicators to contextualized competence assessment. In *11th IEEE International Conference on Advanced Learning Technologies (ICALT), 2011*, pages 277–281. IEEE, 2011. 41, 42, 43, 47, 51, 61
- [16] B. Bell and B. Cowie. The characteristics of formative assessment in science education. *Science education*, 85(5):536–553, 2001. 26
- [17] J. Benlloch-Dualde and S. Blanc-Clavero. Adapting teaching and assessment strategies to enhance competence-based learning in the framework of the european convergence process. In *37th Annual Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07.*, pages S3B–1. IEEE, 2007. 40, 41, 46, 60
- [18] A. Berns, A. Gonzalez-Pardo, and D. Camacho. Game-like language learning in 3-D virtual environments. *Computers & Education*, 60(1):210–220, 2013. 102
- [19] A. Berns, M. Palomo-Duarte, J. M. Doderó, and C. Valero-Franco. Using a 3D online game to assess students' foreign language acquisition and communicative competence. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8095 LNCS:19–31, 2013. 102
- [20] S. Bostock. Student peer assessment. *Learning Technology*, 2000. 59

## BIBLIOGRAFÍA

---

- [21] A. Carreras Marín, Y. Blasco, M. Badia-Miró, M. Bosch Príncipe, I. Morillo, G. Cairó i Céspedes, and D. Casares Vidal. The promotion and assessment of generic skills from interdisciplinary teaching teams. *EDULEARN13 Proceedings*, pages 201–207, 2013. 36, 37, 38, 48, 51, 58
- [22] Y. Chang, T. Eklund, J. I. Kantola, and H. Vanharanta. International creative tension study of university students in south korea and finland. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 19(6):528–543, 2009. 35, 37, 38, 47
- [23] N. Charlier, O. Michela, B. Remmele, and N. Whitton. Not just for children: game-based learning for older adults. In *6th European Conference on Games Based Learning, Cork, Ireland*, pages 102–108, 2012. 42
- [24] M. A. Chatti, V. Lukarov, H. Thijs, A. Muslim, A. M. F. Yousef, U. Wahid, C. Greven, A. Chakrabarti, and U. Schroeder. Learning analytics: Challenges and future research directions. *E-Learning & Education*, 10(1), 2014. 4
- [25] H. Chebil, J. Girardot, and C. Courtin. An ontology-based approach for sharing and analyzing learning trace corpora. In *Proceedings - IEEE 6th International Conference on Semantic Computing, ICSC 2012*, pages 101–108, 2012. 4, 14
- [26] C.-h. Chen. The implementation and evaluation of a mobile self-and peer-assessment system. *Computers & Education*, 55(1):229–236, 2010. 21
- [27] R. Cole, S. Purao, M. Rossi, and M. Sein. Being proactive: where action research meets design research. In *Proceedings - International Conference on Information System (ICIS), 2005*, page 27, 2005. 74
- [28] A. Collins, D. Joseph, and K. Bielaczyc. Design research: Theoretical and methodological issues. *The Journal of the learning sciences*, 13(1):15–42, 2004. 5



- [29] B. Communiqué. Making the most of our potential: Consolidating the european higher education area. In *2012 Ministerial Conference and 3rd Bologna Policy Forum Documents*, Bucharest, 2012. 2
- [30] M. A. Conde, F. J. García-Peñalvo, D.-A. Gómez-Aguilar, and R. Theron. Exploring software engineering subjects by using visual learning analytics techniques. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 10(4):242–252, 2015. 4
- [31] J. Cruz-Benito, R. Therón, F. J. García-Peñalvo, and E. P. Lucas. Discovering usage behaviors and engagement in an educational virtual world. *Computers in Human Behavior*, 47:18–25, 2015. 102
- [32] D. D. Curtis. The assessment of generic skills. *Generic skills in vocational education and training: Research findings*, pages 136–56, 2004. 1
- [33] B. De La Harpe, A. Radloff, and J. Wyber. Quality and generic (professional) skills. *Quality in Higher Education*, 6(3):231–243, 2000. 1
- [34] I. de Los Rios, A. Cazorla, J. M. Díaz-Puente, and J. L. Yagüe. Project-based learning in engineering higher education: two decades of teaching competences in real environments. *Procedia-Social and Behavioral Sciences*, 2(2):1368–1378, 2010. 2
- [35] H. Dixon. What is teamwork? *The Canadian Journal of Medical Radiation Technology/CAMRT*, 23(2):79–80, 1992. 25, 223
- [36] D. Djaouti, J. Alvarez, and J.-P. Jessel. Classifying serious games: the g/p/s model. *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches*, pages 118–136, 2011. 30, 51
- [37] M. P. Duarte, I. M. Bulo, E. J. R. Posada, and F. P. Lozano. Wikis in teaching: An experiment with wikiphaskell and statmediawiki. *RUSC. Universities and Knowledge Society Journal*, 9(1):65–85, 2012. 79

## BIBLIOGRAFÍA

---

- [38] L. Dunn, C. Morgan, and S. Parry. Seeking quality in criterion referenced assessment. In *Learning Communities and Assessment Cultures Conference organised by the EARLI Special Interest Group on Assessment and Evaluation, University of Northumbria, 28-30 August 2002, 2002*. 27
- [39] I. Elgort, A. G. Smith, and J. Toland. Is wiki an effective platform for group course work? *Australasian Journal of Educational Technology*, 24(2):195–210, 2008. 78
- [40] A. P. Ellis, B. S. Bell, R. E. Ployhart, J. R. Hollenbeck, and D. R. Ilgen. An evaluation of generic teamwork skills training with action teams: effects on cognitive and skill-based outcomes. *Personnel psychology*, 58(3):641–672, 2005. 1
- [41] C. European. The Lifelong Learning Programme. In *Decision No 1720/2006/EC of the European Parliament and of the Council of 15 November 2006 establishing an action programme in the field of lifelong learning*, chapter 1, pages 327/48 – 327/49. Official Journal of the European Union, nov 2006. 3
- [42] M. Eysholdt and H. Behrens. Xtext: implement your language faster than the quick and dirty way. In *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, pages 307–309. ACM, 2010. 91
- [43] B. Fernández-Manjón, P. Moreno-Ger, M. Freire, and I. Martínez-Ortiz. *Juegos Serios - Informe CODDII - Conferencia de Decanos y Directores de Ingeniería Informática*. Facultad de Informática, Universidad Complutense de Madrid, 2016. 28, 30
- [44] P. Ficapal-Cusí and J. Boada-Grau. e-learning and team-based learning. practical experience in virtual teams. *Procedia-Social and Behavioral Sciences*, 196:69–74, 2015. 35, 36, 37, 38, 47
- [45] Á. Fidalgo-Blanco, M. L. Sein-Echaluce, F. J. García-Peñalvo, and M. Á. Conde. Using learning analytics to improve teamwork assessment. *Computers in Human Behavior*, 47:149–156, 2015. 42, 43, 45, 48, 51, 54, 62

- 
- [46] A. Fisher. *Critical thinking: An introduction*. Cambridge University Press, 2011. 25, 223
- [47] B. Florian, C. Glahn, H. Drachsler, M. Specht, and R. Fabregat Gesa. Activity-based learner-models for learner monitoring and recommendations in moodle. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6964 LNCS:111–124, 2011. 4
- [48] M. Fowler. *Domain-specific languages*. Pearson Education, 2010. 76
- [49] D. R. Garrison, T. Anderson, and W. Archer. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The internet and higher education*, 2(2):87–105, 1999. 94
- [50] S. M. Gass. *Second language acquisition: An introductory course*. Routledge, 2013. 25, 223
- [51] F. Gates. *Teachers Know Best. Making Data Work For Teachers and Students*. Bill & Melinda Gates Foundation, 2015. 14
- [52] J. González and R. Wagenaar. Tuning Educational Structures in Europe II. Universitites’ contribution to the Bologna Process. *University of Deusto*, 2005. 3, 24, 25, 223
- [53] M. Guenaga, S. Arranz, I. Rubio Florido, E. Aguilar, A. Ortiz de Guinea, A. Rayon, M. J. Bezanilla, and I. Menchaca. Serious games for the development of employment oriented competences. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 8(4):176–183, 2013. 41, 42, 43, 48, 61
- [54] M. Guitert, T. Romeu, and M. Pérez-Mateo. Competencias TIC y trabajo en equipo en entornos virtuales. *RUSC. Universities and Knowledge Society Journal*, 4(1):1–12, 2007. 49
- [55] J. T. Gulikers, T. J. Bastiaens, and P. A. Kirschner. A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3):67–86, 2004. 26

## BIBLIOGRAFÍA

---

- [56] A. R. Hevner. Interview with Alan R. Hevner on “Design Science”. *Business & Information Systems Engineering (BISE)*, 1(1):126–129, 2009. 5
- [57] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design science in information systems research. *MIS Quarterly*, 28(1):75–105, Mar. 2004. 19
- [58] K. F. Hew and W. S. Cheung. Use of three-dimensional (3-d) immersive virtual worlds in k-12 and higher education settings: A review of the research. *British journal of educational technology*, 41(1):33–55, 2010. 102
- [59] G. Hughes. Towards a personal best: a case for introducing ipsative assessment in higher education. *Studies in Higher Education*, 36(3):353–367, 2011. 27
- [60] A. Huhta. Diagnostic and formative assessment. *The handbook of educational linguistics*, 33:469–483, 2008. 26
- [61] J. L. C. Izquierdo and J. Cabot. Community-driven language development. In *ICSE Workshop on Modeling in Software Engineering (MISE), 2012*, pages 29–35. IEEE, 2012. 73
- [62] B. Kitchenham and C. Stuart. Guidelines for performing systematic literature reviews in software engineering. In *Technical report, Ver. 2.3 EBSE Technical Report*. EBSE, 2007. 15
- [63] R. Lacuesta, G. Palacios, and L. Fernández. Active learning through problem based learning methodology in engineering education. In *39th IEEE Frontiers in Education Conference, 2009. FIE'09.*, pages 1–6. IEEE, 2009. 30, 39, 40, 41, 46, 51, 60
- [64] A. Lasa, I. Txurruka, E. Simón, and J. Miranda. Problem based learning implementation in the degree of human nutrition and dietetics. *Proceedings of the International Conference of Education, Research and Innovation (ICE-RI) 2013*, pages 1687–1692, 2013. 35, 36, 37, 38, 40, 47, 60
- [65] D. Laurillard. *Teaching as a design science. Building pedagogical patterns for learning and technology*. Routledge London, 2012. 5, 6

- [66] C. S. E. Lidz. *Dynamic assessment: An interactional approach to evaluating learning potential*. Guilford Press, 1987. 27
- [67] H. Malehorn. Ten measures better than grading. *The Clearing House*, 67(6):323–324, 1994. 34, 58
- [68] M. Martín-Briceño and S. Prashar. Acquired Skills With The Implementation Of New Evaluation Methods At University Rey Juan Carlos. *Proceedings of the International Conference of Education, Research and Innovation (ICERI) 2013*, pages 4875–4878, 2013. 40, 46
- [69] J. E. P. Martinez, J. García Martín, and A. S. Alonso. Teamwork competence and academic motivation in computer science engineering studies. In *IEEE Global Engineering Education Conference (EDUCON), 2014*, pages 778–783. IEEE, 2014. 35, 37, 38, 48
- [70] A. Masip-Álvarez, C. Hervada-Sala, T. Pàmies-Gómez, A. Arias-Pujol, C. Jaen-Fernandez, C. Rodriguez-Sorigue, D. Romero-Duran, F. Nejjar-Akhi-Elarab, M. Alvarez-del Castillo, M. Roca-Lefler, et al. Self-video recording for the integration and assessment of generic competencies. In *IEEE Global Engineering Education Conference (EDUCON), 2013*, pages 436–441. IEEE, 2013. 36, 37, 38, 48
- [71] G. Mason, G. Williams, and S. Cranmer. Employability skills initiatives in higher education: what effects do they have on graduate labour market outcomes? *Education Economics*, 17(1):1–30, 2009. 1
- [72] V. Midoro. *Guidelines on adaptation of the UNESCO ICT Competency Framework for Teachers*. UNESCO Institute for Information Technologies in Education, 2013. 4
- [73] P. Moreno-Ger, I. Martinez-Ortiz, M. Freire, B. Manero, and B. Fernandez-Manjon. Serious games: A journey from research to application. In *IEEE Frontiers in Education Conference (FIE), 2014*, pages 1–4. IEEE, 2014. 102

## BIBLIOGRAFÍA

---

- [74] T. G. Muñoz. El cuestionario como instrumento de investigación/evaluación. [http://cvonline.uaeh.edu.mx/Cursos/Maestria/MTE/Gen02/seminario\\_de\\_tesis/Unidad\\_4\\_anterior/Lect\\_El\\_Cuestionario.pdf](http://cvonline.uaeh.edu.mx/Cursos/Maestria/MTE/Gen02/seminario_de_tesis/Unidad_4_anterior/Lect_El_Cuestionario.pdf), 2003. Accessed: 2016-05-01. 29
- [75] B. J. Oates. *Researching information systems and computing*. Sage, 2005. 11, 112, 135, 145
- [76] B. Oliver. Graduate attributes as a focus for institution-wide curriculum renewal: innovations and challenges. *Higher Education Research & Development*, 32(3):450–463, 2013. 37, 38, 47
- [77] J. Ortega Valiente and J. Reinoso Peinado. New educational approach based on the use of wiki platforms in university environments. In *7th International Conference on Next Generation Web Services Practices (NWeSP), 2011*, pages 280–284. IEEE, 2011. 79
- [78] M. Palomo-Duarte, J. M. Dodero, A. García-Domínguez, P. Neira-Ayuso, N. Sales-Montes, I. Medina-Bulo, F. Palomo-Lozano, C. Castro-Cabrera, E. J. Rodríguez-Posada, and A. Balderas. Scalability of assessments of wiki-based learning experiences in higher education. *Computers in Human Behavior*, 31:638–650, 2014. 109, 136, 145
- [79] M. Palomo-Duarte, J. M. Dodero, I. Medina-Bulo, E. J. Rodríguez-Posada, and I. Ruiz-Rube. Assessment of collaborative learning experiences by graphical analysis of wiki contributions. *Interactive Learning Environments*, 22(4):444–466, 2014. 79
- [80] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering*, volume 17, pages 1–10. British Computer Society, 2008. 15, 19
- [81] N. Piedra, J. Chicaiza, J. López, A. Romero, and E. Tovar. Measuring collaboration and creativity skills through rubrics: Experience from UTPL collaborative social networks course. In *IEEE Education Engineering (EDU-CON), 2010*, pages 1511–1516. IEEE, 2010. 37, 38, 47

- [82] V. Podgorelec and S. Kuhar. Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments. *Elektronika ir Elektrotechnika*, 114(8):111–116, 2011. 14
- [83] QAA. *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education*. The Quality Assurance Agency for Higher Education, 2006. 27
- [84] R. A. Rashid, R. Abdullah, A. Zaharim, H. A. Ghulman, M. S. Masodi, J. Mauri, A. Zaharim, A. Kolyshkin, M. Hatziprokopiou, A. Lazakidou, et al. Engineering students performance evaluation of generic skills measurement: ESPEGS model. In *Proceedings of the 5th WSEAS International Conference on Engineering Education (EE'08)*, volume 5, pages 377–383. World Scientific and Engineering Academy and Society (WSEAS), 2008. 40, 47
- [85] A. Rayon Jerez, M. Guenaga, and A. Núñez. A web platform for the assessment of competences in mobile learning contexts. In *IEEE Global Engineering Education Conference (EDUCON), 2014*, pages 321–329. IEEE, 2014. 42, 43, 44, 46, 51, 54, 62
- [86] C. Redecker and O. Johannessen. Changing Assessment — Towards a New Assessment Paradigm Using ICT. *European Journal of Education*, 48(1):79–96, 2013. 22
- [87] P. Reimann. Design-based research. In *Methodological Choice and Design*, pages 37–50. Springer, 2011. 6
- [88] M. R. Renau and J. U. Viciado. Teaching and learning through projects using the ict: Practice of the english writing through business documents. *Proceedings of the International Conference of Education, Research and Innovation (ICERI) 2010*, pages 4700–4705, 2010. 37, 38, 48
- [89] S. Rodriguez-Donaire, B. A. García, and S. O. Del Olmo. e-portfolio: a tool to assess university students' skills. In *9th International Conference on Information Technology Based Higher Education and Training (ITHET), 2010*, pages 114–124. IEEE, 2010. 40, 47

## BIBLIOGRAFÍA

---

- [90] A. Rojas-Vargas. *La prueba escrita*. Ministerio de Educación Pública de Costa Rica. Dirección de desarrollo curricular. Departamento de Evaluación de los Aprendizajes, 2008. 28
- [91] I. Ruiz-Rube, J. M. Dodero, and R. Colomo-Palacios. A framework for software process deployment and evaluation. *Information and Software Technology*, 59:205–221, 2015. 135, 145
- [92] C. Ruizacárate Varela, M. J. Garcia Garcia, C. Gonzalez Garcia, and J. L. Casado Sanchez. Soft skills: A comparative analysis between online and classroom teaching. In *Proceedings of the International Conference on Advanced Education Technology and Management Science (AETMS2013)*, pages 359–366. DEStech Publications, 2013. 37, 38, 48
- [93] J. Serrano-Guerrero, F. P. Romero, and J. A. Olivas. Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences. *Information Sciences*, 248:114–129, 2013. 40, 41, 47, 60
- [94] A. Sevilla-Pavón, A. Martínez-Sáez, and A. Gimeno-Sanz. Assessment of competences in designing online preparatory materials for the cambridge first certificate in english examination. *Procedia-Social and Behavioral Sciences*, 34:207–211, 2012. 36, 37, 38, 46, 60
- [95] A. I. Starcic. Sustaining teacher’s professional development and training through web-based communities of practice. In *International Symposium on Applications and the Internet, 2008. SAINT 2008.*, pages 317–320. IEEE, 2008. 35, 39, 40, 48
- [96] J. Strijbos, N. Engels, and K. Struyven. Criteria and standards of generic competences at bachelor degree level: A review study. *Educational Research Review*, 14:18–32, 2015. 3, 46, 49
- [97] J.-J. Torres-Gordillo and V.-H. Perera-Rodríguez. La rúbrica como instrumento pedagógico para la tutorización y evaluación de los aprendizajes en el foro online en educación superior. *Pixel-Bit: Revista de medios y educación*, 36:141–149, 2010. 29



- [98] G. Trentin. Using a wiki to evaluate individual contribution to a collaborative learning project. *Journal of Computer Assisted Learning*, 25(1):43–55, 2009. 79
- [99] C. Tünnermann Bernheim. La educación permanente y su impacto en la educación superior. *Revista Iberoamericana de Educación Superior (RIES), México, issue - unam /Universia*, 1(1):120–133, 2010. 25, 223
- [100] R. W. Tyler, R. M. Gagné, M. Scriven, et al. *Perspectives of curriculum evaluation*, volume 1. Rand McNally Chicago, 1967. 26
- [101] C. Vizcarro Guarch, P. Martin Espinosa, R. Cobos, J. E. Pérez, E. Tovar Caro, G. Blanco Viejo, A. Bermudez Marin, and J. Reyes Ruiz Gallardo. Assessment of problem solving in computing studies. In *IEEE Frontiers in Education Conference, 2013*, pages 999–1003. IEEE, 2013. 40, 46
- [102] T. Ward and S. Christophe. Developing entrepreneurial accounting and finance competency using the elleiec virtual centre for enterprise. In *Proceedings of the 22nd European Association For Education In Electrical And Information Engineering Annual Conference (EAEEIE)*, pages 1–5. IEEE, 2011. 29, 39, 40, 47
- [103] R. Wieringa, N. Maiden, N. Mead, and C. Rolland. Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. *Requirements Engineering*, 11(1):102–107, Dec. 2005. 19
- [104] J. S. Wiggins. *Paradigms of personality assessment*. Guilford Press, 2003. 30
- [105] F. Yang, F. W. Li, and R. W. Lau. A fine-grained outcome-based learning path model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(2):235–245, 2014. 39, 40, 46
- [106] J. G. Zabala. *El espacio europeo de educación superior, un reto para la universidad: competencias, tareas y evaluación, los ejes del currículum universitario*. Barcelona : Editorial Octaedro, S.L., 2005. 3

Este tesis doctoral fue finalizada en Jerez de la Frontera el 17 de junio de 2016

*Esta página se ha dejado en blanco intencionadamente.*