



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, abril 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

The following web-page address contains up to date information about this dissertation and related topics:

<http://paginaspersonales.deusto.es/Name/>

Impreso en Jerez de la Frontera

Primera edición, abril 2016

Dedicatoria.

Dedico este trabajo a ...

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut ultrices egestas nunc, venenatis rhoncus elit fermentum non. Pellentesque gravida nulla vitae ipsum lobortis ullamcorper. Ut adipiscing, tellus in egestas mattis, enim metus pretium erat, ac tempor dolor neque placerat nulla. Nullam nec ligula eu ipsum pharetra semper a in magna. Integer ut tortor quis nisi fringilla euismod eu ac ipsum. Pellentesque sodales consectetur erat eget rutrum. Proin ornare dolor ut arcu aliquet vestibulum. Pellentesque laoreet tincidunt sem eget semper.

Integer interdum mattis magna ullamcorper tristique. Nullam commodo nulla eget ipsum vulputate tincidunt auctor leo aliquet. Fusce euismod sagittis ante, eu vulputate eros dictum at. Cras non euismod nunc. Nullam velit diam, consectetur sed eleifend vitae, blandit at arcu. Maecenas ut urna nec turpis lobortis commodo. Aliquam aliquet turpis id massa viverra id sollicitudin est cursus. Sed a tortor non mauris cursus imperdiet.

Integer fermentum rutrum urna at vestibulum. Vivamus ullamcorper erat in sapien dignissim pellentesque. Integer convallis fringilla dictum. In bibendum lectus eu nulla pretium volutpat. Morbi hendrerit fringilla tortor, sed gravida neque lacinia a. In risus magna, hendrerit vitae cursus ac, vehicula at eros. Aenean quis ipsum sit amet leo vestibulum cursus.

Cras placerat mattis dui quis vehicula. Nulla sit amet metus nibh, at auctor enim. Quisque congue ultricies sapien in suscipit. Fusce vitae placerat ante. Praesent aliquet urna ac elit consequat nec mattis

augue faucibus. Nunc et sapien vel felis mollis sodales. Aenean molestie nulla vestibulum nisi fringilla vel euismod dolor tristique. Aenean fermentum, dolor eget tincidunt faucibus, risus lorem feugiat elit, sagittis malesuada eros ligula in odio. Pellentesque ac libero lobortis justo bibendum laoreet. Cras egestas lorem eget ligula dignissim sollicitudin. Vestibulum sit amet augue ultrices erat faucibus vestibulum. Aenean tincidunt faucibus leo, nec auctor diam bibendum a. Sed varius, mauris in pellentesque scelerisque, nisl ligula viverra erat, in eleifend tellus enim ac magna. Pellentesque quis est risus. Cras mollis feugiat auctor. Proin ac eros vitae nulla gravida varius.

Morbi at augue sapien. Duis tempus quam vitae velit interdum ultricies. Vivamus laoreet lacinia elit sit amet vehicula. Ut congue diam ac magna hendrerit sed fermentum justo lacinia. Curabitur vel odio neque, quis consequat mi. Proin lobortis justo quis enim fermentum accumsan sagittis ipsum imperdiet. Proin sem felis, laoreet placerat egestas id, fringilla id mauris. Pellentesque a nisi sit amet leo consectetur gravida nec et dui. Curabitur quis hendrerit augue. Etiam sed dui nec tortor convallis fringilla. Proin tempor mattis diam nec egestas. Quisque condimentum elementum lacus ac porta. Vivamus congue, odio eu ullamcorper elementum, leo turpis tempus sem, at condimentum dolor quam eu nunc. Pellentesque eget risus ac velit aliquam sollicitudin sed et ipsum.

Resumen

La evaluación de competencias genéricas es una labor que tienen que realizar los docentes de todos los niveles educativos en sus asignaturas.

En este trabajo se ha desarrollado un método de evaluación de competencias genéricas basado en el diseño (DBA). Este método se basa en el diseño de evaluaciones mediante la utilización de indicadores de la actividad desempeñada por los estudiantes en los entornos de aprendizaje virtual. Este método surge de la oportunidad que se presenta de aprovechar la cantidad de información generada por los estudiantes en los entornos virtuales.

Para poner en práctica el método DBA se han desarrollado dos DSL que lo implementan. En primer lugar se desarrolló SASQL, un DSL para obtener indicadores de la actividad de los estudiantes del VLE y EvalCourse, el software que interpreta las consultas escritas en SASQL. En segundo lugar se desarrolló VWQL, DSL que en este caso obtiene indicadores de la actividad de los estudiantes de los mundos virtuales y EvalSim, el software que interpreta las consultas escritas en VWQL.

Se han realizado varios estudios de caso en los que se han planificado actividades en el campus virtual para después recopilar los rastros de interacción generados por los estudiantes y aplicarlos a la evaluación de varias competencias genéricas.

La evaluación de la metodología se ha realizado mediante encuestas ...

Acknowledgements

Quiero dar las gracias a todos mis compañeros y en especial a ...

Gracias,

Antonio

abril 2016

Índice general

Índice de figuras	xv
Índice de tablas	xix
1 Introducción	1
1.1 Motivación	3
1.2 Contexto	7
1.3 Objetivos y preguntas de investigación	10
1.3.1 Preguntas de investigación	10
1.3.2 Objetivos	11
1.3.3 Hipótesis de investigación	11
1.4 Estrategia de investigación	13
1.4.1 Diseño y creación	13
2 Estado del Arte	15
2.1 Preguntas de investigación	16
2.2 Metodología	17
2.2.1 Criterios de selección	19
2.2.2 Esquema para la extracción de datos	20
2.2.2.1 Tipo de investigación	20
2.2.2.2 Tipo de contribución	21
2.2.2.3 Ámbito de aplicación de la investigación	22
2.2.3 Visualización y análisis de los datos	23
2.3 Resultados	23

ÍNDICE GENERAL

2.3.1	Localización de la literatura	24
2.3.2	Extracción de los datos	25
2.3.2.1	Competencias genéricas	27
2.3.2.2	Métodos	29
2.3.2.3	Técnicas	30
2.3.2.4	Instrumentos de evaluación	32
2.3.3	Categorización del estudio	33
2.3.4	Esquema de clasificación	37
2.3.4.1	Autoevaluación o evaluación entre iguales.	37
2.3.4.2	Evaluación del profesor	41
2.3.4.3	Evaluación semiautomática	44
2.4	Respuestas a las preguntas de investigación	52
2.5	Conclusiones	58
3	Resumen de problemas encontrados	61
3.1	Evaluación asistida	61
3.1.1	Autoevaluación o evaluación entre iguales	62
3.1.2	Evaluación del profesor	63
3.2	Evaluación semiautomática	63
4	Método para la evaluación de competencias genéricas	67
4.1	Introducción	67
4.2	Método: design-based assessment (DBA)	68
4.2.1	Contexto	68
4.2.2	Descripción del método	69
4.3	Características, requisitos y herramienta	73
4.3.1	Características	73
4.3.2	Requisitos	76
4.3.3	Herramienta	78
5	Resultados de la investigación	81
5.1	Introducción	81
5.2	Antecedentes	82
5.2.1	Introducción	82

ÍNDICE GENERAL

5.2.2	AssessMediaWiki (AMW)	84
5.2.2.1	Ejemplo de uso	85
5.2.2.2	Conclusiones	94
5.3	Lenguajes de dominio	94
5.3.1	EvalCourse (EVC) y SASQL	94
5.3.1.1	Descripción	95
5.3.1.2	Ejemplo	97
5.3.2	EvalSim (EVS) y VWQL	104
5.3.2.1	Descripción	104
5.3.2.2	Ejemplo	106
5.4	Estudios de caso	110
5.4.1	Estudio de caso CS01	110
5.4.2	Estudio de caso CS02	111
5.4.3	Estudio de caso CS03	111
5.4.4	Estudio de caso CS04	111
5.4.5	Estudio de caso CS05	112
5.4.6	Estudio de caso CS06	112
5.5	Cuestionarios de evaluación	112
5.5.1	Objetivo	113
5.5.2	Participantes	114
5.5.2.1	Curso wiki innovación	114
5.5.2.2	Taller Aulablog	115
5.5.2.3	Actuación avalada	115
5.5.2.4	MediaWiki España	115
5.5.2.5	Jornadas de Innovación Docente de la Universidad de Cádiz	116
5.5.3	Resultados del cuestionario del método y el DSL	116
5.5.3.1	Perfil de los participantes	116
5.5.3.2	Evaluación del método DBA	117
5.5.3.2.1	Características del método DBA	118
5.5.3.3	Evaluación del DSL	119
5.5.3.3.1	Características del DSL	120
5.5.3.4	Evaluación de la sintaxis del DSL (consulta tipo)	121

ÍNDICE GENERAL

5.5.3.5	Evaluación de los resultados	122
5.5.4	Resultados del cuestionario de indicadores	123
5.5.4.1	Perfil de los participantes	124
5.5.4.2	Evaluación de competencias genéricas	125
5.5.4.2.1	Independencia de las respuestas	126
5.5.4.2.2	Justificaciones	127
5.5.5	Resultados de la encuesta complementaria de indicadores .	128
5.5.6	Discusión y análisis	129
5.6	Contribuciones	132
6	Conclusiones y trabajo futuro	135
6.1	Conclusiones	135
6.2	Trabajo futuro	137
A	Cuestionario	139
A.1	Cuestionario	140
A.2	Resultados	145
A.2.1	Poblaciones	145
A.2.1.1	Curso innovación docente	145
A.2.1.2	Taller Aulablog	145
A.2.1.3	Actuación avalada	146
A.2.1.4	MediaWiki	146
A.2.2	Participantes	146
A.2.3	Indicadores de participación y de contribución	147
A.2.4	Caso de evaluación de competencias genéricas	148
A.2.4.1	Conocimientos de programación	148
A.2.4.2	Perfil	153
A.2.4.3	Comentarios	157
B	Cuestionario complementario actuación avalada	167
B.1	Cuestionario	168
B.2	Resultados	171

C	Evaluación del método de evaluación (DBA) y del lenguaje para diseñar evaluaciones (SASQL)	177
C.1	Cuestionario	178
C.2	Resultados	183
C.2.1	Participantes	183
C.2.2	Evaluación del método DBA	183
C.2.2.1	Características del método DBA	184
C.2.3	Evaluación del DSL	191
C.2.3.1	Características del DSL	192
C.2.4	Evaluación de la sintaxis del DSL (consulta tipo)	199
C.2.5	Evaluación de los resultados	200
	Bibliografía	201

Índice de figuras

2.1	Distribución de las publicaciones por años	26
2.2	Ámbito de trabajos distribuidos según tipo de investigación y según tipo de contribución.	36
2.3	Competencias genéricas evaluadas con cada método (1 de 2). . . .	55
2.4	Competencias genéricas evaluadas con cada método (2 de 2). . . .	56
4.1	Diagrama del ciclo de contraste de hipótesis	69
4.2	Interacción en el foro en un periodo de tiempo.	73
5.1	Rúbrica de AMW	84
5.2	Ejemplo de retroalimentación formativa y la contribución de wiki evaluada	85
5.3	Ejemplo de flujo de trabajo para la evaluación cualitativa del wiki utilizando AMW	86
5.4	Ciclo de contraste de hipótesis para la evaluación en wikis	91
5.5	Metamodelo de SASQL	96
5.6	Ciclo de contraste de hipótesis utilizando EvalCourse	99
5.7	Interacción en el foro en un periodo de tiempo.	100
5.8	Ciclo de contraste de hipótesis con EvalSim	106
5.9	Distribución de los participantes por rama y sexo	116
5.10	Respuestas de los encuestados a la pregunta sobre el método DBA .	117
5.11	Distribución de los participantes en el cuestionario	124
6.1	Grafo resultante de aplicar técnicas de SNA a un wiki de Moodle .	138

ÍNDICE DE FIGURAS

A.1	Participación de los estudiantes en la página del wiki	141
A.2	Contribuciones de los estudiantes a la página del wiki	142
A.3	Evolución del contenido de una página del wiki	143
A.4	Utilizaría el indicador para evaluar el trabajo en equipo	149
A.5	Utilizaría el indicador para evaluar la planificación y gestión de tiempo	151
A.6	Utilizaría el indicador para evaluar el liderazgo	152
A.7	Utilizaría el indicador para evaluar el trabajo en equipo	153
A.8	Utilizaría el indicador para evaluar la planificación y gestión de tiempo	155
A.9	Utilizaría el indicador para evaluar el liderazgo	156
B.1	Docentes que considerarían los indicadores de accesos para evaluar cada competencia genérica	171
B.2	Docentes que considerarían los indicadores de los foros para eva- luar cada competencia genérica	173
B.3	Docentes que considerarían los indicadores de los wikis para eva- luar cada competencia genérica	174
B.4	Docentes que considerarían los indicadores de las actividades para evaluar cada competencia genérica	175
C.1	Resultados consulta	181
C.2	Resumen de respuestas para la característica de la objetividad . . .	186
C.3	Resumen de respuestas para la característica de la objetividad por rama	186
C.4	Resumen de respuestas para la característica de la adaptabilidad .	187
C.5	Resumen de respuestas para la característica de la adaptabilidad por rama	187
C.6	Resumen de respuestas para la característica de la sistematicidad .	188
C.7	Resumen de respuestas para la característica de la sistematicidad por rama	188
C.8	Resumen de respuestas para la característica de la flexibilidad . .	189
C.9	Resumen de respuestas para la característica de la flexibilidad por rama	189

ÍNDICE DE FIGURAS

C.10 Resumen de respuestas para la característica de la fiabilidad	190
C.11 Resumen de respuestas para la característica de la fiabilidad por rama	190
C.12 Resumen de respuestas para la característica de la facilidad de apren- dizaje	193
C.13 Resumen de respuestas para la característica de la facilidad de apren- dizaje por rama	193
C.14 Resumen de respuestas para la característica de la eficacia	194
C.15 Resumen de respuestas para la característica de la eficacia por rama	194
C.16 Resumen de respuestas para la característica del ahorro del tiempo	195
C.17 Resumen de respuestas para la característica del ahorro del tiempo	195
C.18 Resumen de respuestas para la característica de la escalabilidad . .	196
C.19 Resumen de respuestas para la característica de la escalabilidad por rama	196
C.20 Resumen de respuestas para la característica de la capacidad de reutilización	197
C.21 Resumen de respuestas para la característica de la capacidad de reutilización por rama	197
C.22 Resumen de respuestas para la característica de la fiabilidad	198
C.23 Resumen de respuestas para la característica de la fiabilidad por rama	198

Índice de tablas

2.1	Resumen de búsqueda de bibliografía	19
2.2	Bibliotecas digitales utilizadas, palabras de búsqueda utilizadas en cada uno y número de resultados obtenidos	24
2.3	Clasificación de trabajos una vez aplicados los criterios de selección y exclusión	25
2.4	Cantidad de trabajos publicados cada año	26
2.5	Competencias genéricas	28
2.6	Distribución de publicaciones por tratamiento del problema	34
2.7	Competencias genéricas evaluadas mediante autoevaluación y evaluación entre iguales	40
2.8	Instrumentos de evaluación y métodos correspondientes a los trabajos de la autoevaluación y evaluación entre iguales	41
2.9	Competencias evaluadas directamente por el profesor	43
2.10	Instrumentos de evaluación y métodos correspondientes a los trabajos en los que la evaluación es realizada por los profesores . . .	43
2.11	Competencias evaluadas de forma semiautomática	46
2.12	Instrumentos de evaluación y métodos correspondientes a los trabajos de evaluación semiautomática	46
2.13	Distribución de publicaciones por tratamiento del problema	51
2.14	Número de trabajos que evalúan cada competencia genérica	53
2.15	Competencias evaluadas, métodos aplicados y técnicas seguidas para evaluar competencias a partir de los registros de actividad de los entornos de aprendizaje	57

ÍNDICE DE TABLAS

4.1	Información sobre la participación en el foro de los estudiantes en un periodo concreto de tiempo	72
5.1	Resumen de las competencias evaluadas para cada tipo de indicador	93
5.2	Información sobre la participación en el foro de los estudiantes en un periodo concreto de tiempo	99
5.3	Información sobre las respuestas de una sola palabra dadas por los estudiantes en el role-play	107
5.4	Información sobre las palabras por turnos utilizadas por los estudiantes en el role-play	108
5.5	Resumen de publicaciones de los estudios de caso	110
5.6	Resumen de distribución de participantes por rama y sexo	117
5.7	Consideración de idoneidad del método DBA por ramas	118
5.8	Evaluación de las características del método DBA	119
5.9	Consideración de idoneidad del DSL por ramas	120
5.10	Evaluación de las características del DSL	121
5.11	Consideración de la consulta como entendible o no por ramas . . .	122
5.12	Consideración de los resultados como entendibles o no por ramas .	123
5.13	Perfil de los participantes que participaron en el cuestionario . . .	125
5.14	Resumen de la validez de cada indicador según los participantes en el cuestionario para evaluar cada competencia genérica	125
5.15	Cuadro de indicadores considerados para cada competencia	129
A.1	Resumen de participantes en el cuestionario	147
A.2	Perfil de los participantes en el cuestionario	147
A.3	Respuestas dadas a la consideración como indicadores de actividad	148
A.4	Resumen de la validez dada por los participantes a los indicadores para evaluar competencias genéricas	149
C.1	Selección de características del método DBA	179
C.2	Selección de características del lenguaje SASQL	180
C.3	Resumen de distribución de participantes por rama y sexo	183
C.4	Consideración de idoneidad del método DBA por ramas	184
C.5	Evaluación de las características del método DBA	185

ÍNDICE DE TABLAS

C.6	Consideración de idoneidad del DSL por ramas	191
C.7	Evaluación de las características del DSL	192
C.8	Consideración de la consulta como entendible o no por ramas . . .	199
C.9	Consideración de los resultados como entendibles o no por ramas .	200

Las competencias, las tareas y su evaluación son los pilares en los que se basa el nuevo currículum universitario.

Zabala

CAPÍTULO

1

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 [38, 68]. La investigación educativa muestra que dichas competencias se enseñan con mayor eficacia en el periodo formativo del individuo [31].

La evaluación de competencias genéricas se ha abordado desde diferentes perspectivas. En la revisión de la literatura realizada por Curtis [30] se identificaron cuatro enfoques diferentes: el primero era la evaluación desde el punto de vista del profesor, un enfoque muy eficaz sobre todo a nivel escolar, donde el profesor 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 al profesor un relato muy detallado de sus logros, pero que no está en un formato que sea fácilmente digerible ni comparable para el profesor; el tercer enfoque es la evaluación basada en la experiencia laboral, que parece ser un método útil para producir una evaluación

1. INTRODUCCIÓN

rápida, pero 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 traen 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 [5]. Además, los profesores pueden beneficiarse mucho de los avances tecnológicos para hacer su trabajo más atractivo y ser más eficientes en todas sus actividades cotidianas, entre ellas las tareas de 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 [27], 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 [32]. 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 ¹ 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 [104].

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* [49], apoyado por el Lifelong Learning Programme de la Unión Europea [39], muestran la importancia de utilizar el concepto de competencia como base para los resultados de aprendizaje. Las competencias de aprendizaje 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 [94]. 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. Aunque obviamente sigue siendo muy importante el desarrollo del conocimiento específico de cada área de estudio, es un hecho que el tiempo y la atención también deben dedicarse al desarrollo de las competencias genéricas.

¹<http://www.eees.es/>

1. INTRODUCCIÓN

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 UNESCO, es preciso que los docentes reciban los instrumentos necesarios para alcanzar los objetivos sociales y económicos que constituyen el eje de todo sistema educativo nacional. Para ello, se ha creado el marco de competencias de los docentes en materia de TIC de la UNESCO [69], un conjunto de baremos internacionales 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 cursos virtuales, los wikis o los mundos virtuales son más que habituales como soporte a la docencia presencial, y en algunos casos, como ocurre con los cursos online masivos y abiertos (*MOOCs*, del inglés *Massive Open Online Courses*), es el único punto de encuentro entre el estudiante y el profesor.

En la mayoría de estas herramientas la actividad generada por cada estudiante suele quedar registrada. Entiéndase por *actividad generada* a la recopilación de 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 [23, 44] 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. Las técnicas de *Learning Analytics* facilitan la explotación de este tipo de información [28]. El *learning analytics* es un área de investigación dentro del TEL que está enfocada en el desarrollo de métodos para analizar y detectar patrones en los datos recogidos en los entornos educativos y

aprovecharlos para mejorar el aprendizaje [22]. 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 utilizarán como evidencias de una u otra competencia es tarea de investigadores y profesores. A menos que se pueda considerar que la evaluación de una competencia genérica esté directamente conectada con la evaluación de una actividad específica en la herramienta, su evaluación requiere de la inventiva y originalidad del profesor 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 real del estudiante en dicha competencia a partir de la interacción de este con la herramienta. Ya que tanto los investigadores como los profesores se enfrentan a tareas de diseño, podemos decir que la educación se puede considerar como una *ciencia de diseño* [62].

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 [26].

Para afrontar esta situación surgió una metodología de *investigación del diseño* (DBR, del inglés, design-based research) que no es clásicamente experimental, sino iterativa, progresivamente refinando el diseño inicial basado en la teoría. 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

1. INTRODUCCIÓN

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 [53]

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 profesores en general no investigan, sino que simplemente desarrollan y comparten teorías y explicaciones basadas en su propia experiencia [62]. Además, cuando un profesor 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 las investigación y la práctica en la educación formal [3].

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 [84]. DBR no es una metodología clásicamente experimental, sino iterativa, que se basa en ir refinando progresivamente el diseño inicial basado en la teoría. Según el análisis realizado por Terry Andersen y Julie Shattuck [3], 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 profesores. 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

1. 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 profesores*: por una lado, los profesores 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 principios 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 de diseño prácticos es una parte fundamental del DBR, y pone en desventaja a aquellos 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 profesores 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 profesor, 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. INTRODUCCIÓN

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 de aprendizaje virtual

1.3.1 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:

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?

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 buscan la automatización del proceso para facilitar la labor del docente.

Q2. ¿Qué métodos se utilizan para evaluar competencias genéricas mediante el uso de entornos virtuales?

Una de las contribuciones de esta tesis es un método para la evaluación de competencias genéricas. 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?

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.

1.3.2 Objetivos

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 de aprendizaje virtual, así como una herramienta informática que implemente el método, se obtienen dos objetivos:

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

Basándonos en la metodología DBR se definirá un método que permita evaluar las competencias genéricas de los estudiantes a partir de su actividad en los entornos virtuales.

O2. Definir un DSL (*Lenguaje Específico de Dominio*, del inglés, *Domain-Specific Language*) que permita a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos de aprendizaje virtual

Se definirán un conjunto de herramientas informáticas basadas en un DSL que implementen el método del objetivo O1 y que sea aplicable a diferentes entornos de virtuales de aprendizaje.

1.3.3 Hipótesis de investigación

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

1. INTRODUCCIÓN

H1: El método DBA permite obtener de manera automática indicadores de los entornos de aprendizaje virtual.

Esta hipótesis se corresponde con el objetivo *O1* y persigue conocer si el método propuesto permite obtener indicadores de los entornos de aprendizaje de manera automática.

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

Esta hipótesis se corresponde con el objetivo *O2* y persigue evaluar si la herramienta creada para poner en práctica el método permite a los docentes diseñar y contrastar las estrategias de evaluación a partir de los registros de actividad.

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.

1.4.1 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, ya que es una estrategia que se basa en el desarrollo de un nuevo artefacto o producto tecnológico [72]. Los tipos de artefactos que abarca esta estrategia son constructores (*constructs*), modelos (*models*), métodos (*methods*) e instanciaciones (*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 como 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.

Personally, I think it does help, that it makes a beneficial difference, but the scientific literature on the subject is very messy.

Jeanne Petrek

CAPÍTULO

2

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, además de las competencias específicas que se les supone por la titulación que hayan estudiado. Desde un punto de vista formativo, los profesores deben integrar estas competencias en sus asignaturas, tanto en las clases tradicionales como en los entornos virtuales. Y por supuesto, deben fijar mecanismos no sólo para el desarrollo de estas competencias, sino también para la evaluación de las mismas.

Los *entornos virtuales de aprendizaje* (VLE, del inglés, virtual learning environment) almacenan información de estudiantes, profesores, cursos, tareas, trabajos, etc. Estos elementos se relacionan y configuran para ofrecer al usuario una experiencia de curso virtual. Estos cursos están en auge 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 siempre será mayor. Además, para los estudiantes presentan otras ventajas

2. ESTADO DEL ARTE

fundamentales: en 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 profesores mediante diferentes herramientas (foros, chats, ... etc.) [2].

Pero además de todo lo anterior, un VLE almacena una gran cantidad de información que adecuadamente analizada y presentada podría ser de gran utilidad para los profesores para monitorizar el trabajo de sus estudiantes [79]. 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 profesor, y si lo está, requiere un filtrado para poder ser utilizada [23]. Según el informe *Making Data Work For Teachers and Students*, de la *Fundación Bill & Melinda Gates* [48], el 67% de los profesores 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 profesores 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 VLE podría darnos indicadores sobre cómo actuarían los estudiantes en la vida real y podrían responderse preguntas sobre cómo de eficaz es un estudiante trabajando en equipo, sobre cómo se planifica un estudiante a la hora de realizar sus tareas o sobre si lidera eficazmente un grupo. ¿Podríamos entonces afirmar que los registros de actividad de un VLE pueden utilizarse para evaluar competencias genéricas? Para responder a estas cuestiones y establecer la base teórica sobre la que se sustenta esta tesis doctoral se va a realizar 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

El objetivo principal de esta tesis doctoral 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. Las preguntas habrán de dar respuesta a interrogantes tales como cuáles son las competencias genéricas que se han evaluado haciendo uso de métodos y técnicas informáticas, así como cuáles son esos métodos y si se están usando para este fin los registros de actividad de los entornos virtuales.

Por tanto, partiendo del objetivo principal, se definen las siguientes 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?
- Q2. ¿Qué métodos se utilizan para evaluar competencias genéricas mediante el uso de entornos virtuales?
- Q3. ¿Qué técnicas se utilizan para evaluar competencias genéricas a partir de los registros de actividad de un entorno virtual?

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 publicadas en la metodología propuesta por Kitchenham [59]. Esta metodología describe cómo se deben planificar, ejecutar y presentar los resultados de una revisión de la literatura en ingeniería del software. Para este trabajo se ha utilizado la propuesta de Petersen [77]. 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 realización 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 los 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

SOURCE	SEARCH TERMS	PUBLICATION	RSLT
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

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

2. ESTADO DEL ARTE

- Included: trabajo relacionado con nuestra investigación.
- 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.
- Unsupported Language: trabajo escrito en un lenguaje diferente al inglés o español. La mayoría de los textos son en inglés, por lo que este criterio de descarte apenas es utilizado.
- Duplicated: trabajos cuya contribución principal está recogida en otros trabajos ya incluidos.
- 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 otros repositorios 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.

2.2.2.1 Tipo de investigación

Esta clasificación hace referencia al tipo de trabajo de investigación llevado a cabo por el/los investigador/es. 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 [101] y Hevner [54]. Usamos el primero, ya que es el recomendado en el SMS descrito por Petersen [77].

- 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 del autor.
- Artículos de opinión (*opinion papers*): estos trabajos expresan la opinión personal de alguien acerca de la bondad o viabilidad de una determinada técnica, o cómo se deben realizar las cosas. No se basan en metodologías de trabajo y de investigación relacionadas.
- Trabajos filosóficos (*philosophical papers*): estos trabajos esbozan una nueva forma de ver las cosas existentes, estructurando el campo en forma de una taxonomía o un marco conceptual.

2.2.2.2 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

2. ESTADO DEL ARTE

de la literatura y habiendo seleccionado los artículos, se realiza una clasificación en base a su aportación. El uso de algunos términos puede ser confuso, debido a la interpretación que hace el autor del mismo. Algunos de estos términos son framework, modelo, estrategia, proceso, procedimiento, método o metodología. Nuestra clasificación 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.

2.2.2.3 Ámbito de aplicación de la investigación

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

- **Evaluación del profesor (*teacher assessment*):** el profesor evalúa el desempeño de los estudiantes en una o varias competencias genéricas de manera asistida o semi-asistida por el ordenador.

- 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 artefacto tecnológico. Los investigadores animan a los profesores 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 [24].
- Herramientas de evaluación automática (*automatic assessment tools*): en esta rama se recogen trabajos que automatizan 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 así 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 en un mapa sistemático en la forma de un gráfico de burbujas. En el siguiente capítulo se mostrarán 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 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. ESTADO DEL ARTE

SOURCE	SEARCH TERMS	RESULTS
Web of Science	((“generic competences“ OR “generic skills“) AND assessment)	50
Wiley Online Library	“generic competences“ AND assessment	138
Science Direct	(“generic competences“) AND assessment)	71
IEEE Digital Library (Xplore)	((“generic competences“) AND assessment)	54
TOTAL		313

Tabla 2.2: Bibliotecas digitales utilizadas, palabras de búsqueda utilizadas en cada uno y número de resultados obtenidos

2.3.1 Localización de la literatura

En la tabla 2.2 se muestran las búsquedas realizadas en las bibliotecas digitales más importantes en ciencias de la computación, 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 de búsqueda de este SMS está disponible para su consulta ¹.

En total se recopilaron 313 trabajos para ser revisados. 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 (*off topic*) ya que aunque trataban en su mayoría la evaluación de competencias genéricas, no abordaban su evaluación con apoyo de tecnología, o si lo hacían no se indicaba. Los resultados de esta clasificación pueden verse en la tabla 2.3. A tenor de los resultados ofrecidos, y sobre todo del número de trabajos descartados, el lector puede pensar que se debieron añadir términos computacionales entre los criterios de búsqueda. Sin embargo, el número de términos informáticos que se podrían añadir a las búsquedas es muy amplio y 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,

¹<https://goo.gl/rzTCHN>

2.3 Resultados

CRITERIO	TRABAJOS	PORCENTAJE
Included	30	9,58%
Off Topic	268	85,62%
Unsupported Language	0	0,00%
Duplicated	10	3,20%
Unread	5	1,60%
TOTAL	313	100,00%

Tabla 2.3: Clasificación de trabajos una vez aplicados los criterios de selección y exclusión

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 la tecnología.

2.3.2 Extracción de los 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 bajo el marco de la Comisión Europea (*Generation 3: continuous integrated assessment*) [83], cuando se comienzan a integrar la evaluación en las herramientas de aprendizaje. Entonces conceptos como *Data Mining and analysis*, *Behavioural tracking* and *Learning analytics* comienzan a usarse. Tanto en la tabla 2.4 como en la figura 2.1 puede verse la distribución de la producción de la selección primaria a lo largo de los años. Casi la mayor parte 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 va a concretar para cada trabajo las competencias que evalúa, con qué método, técnica y herramienta lo hace. A continuación se van a describir para cada una de estas características los valores que aparecerán en la revisión para que

2. ESTADO DEL ARTE

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 2.4: Cantidad de trabajos publicados cada año



Figura 2.1: Distribución de las publicaciones por años

queden definidos todos los términos que aparecerán en el capítulo del esquema de clasificación.

2.3.2.1 Competencias genéricas

En la tabla 2.5 se muestran las competencias genéricas que se evalúan en los trabajos seleccionados. 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* [49]. De esta manera, unificábamos la denominación de las competencias, ya que en ocasiones los autores se refieren a las mismas competencias de diferentes maneras.

COMPETENCIA	DESCRIPCIÓN
Análisis	Capacidad de abstracción, análisis y síntesis [49].
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 [97].
Comunicación	Habilidad para comunicarse de manera tanto oral como escrita en la lengua materna [49].
Creatividad	Capacidad para crear nuevas ideas [49].
Cultural	Aprecio y respeto por la diversidad y la multiculturalidad [49].
Emprendimiento	Capacidad para tomar la iniciativa y espíritu de empresa [49].
Gestión de proyectos	Habilidades para diseñar y gestionar proyectos [49].
Habilidades interpersonales	Capacidad de la persona para comunicarse e interactuar con otras personas [49].
Investigación	Capacidad para llevar a cabo la investigación en un nivel apropiado [49].
Liderazgo	Habilidad para motivar a la gente y conducirlos hacia un objetivo común [49].
Pensamiento crítico	Habilidad para interpretar, analizar y evaluar ideas y argumentos [43].
Planificación y gestión del tiempo	Capacidad de planificar y gestionar el tiempo de manera efectiva [49].
Resolución de problemas	Habilidad para identificar, plantear y resolver problemas [49].
Responsabilidad	Capacidad para actuar con responsabilidad social y conciencia cívica [49].
Segundo idioma	Capacidad de los estudiantes para comunicar sus ideas en un segundo idioma [47].
TIC	Habilidades en el uso de las tecnologías de la información y de la comunicación [49].
Toma de decisiones	Capacidad para tomar decisiones razonadas [49].
Trabajo autónomo	Capacidad para trabajar de forma autónoma [49].
Trabajo en equipo	Trabajo realizado por un conjunto de personas en aras de un objetivo común [33].

Tabla 2.5: Competencias genéricas

2.3.2.2 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 profesores y estudiantes para identificar y reaccionar al aprendizaje de los estudiantes con el fin de mejorar dicho aprendizaje mientras éste tiene lugar [15]. 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 si se han alcanzado 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 [98].
- 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 [52].
- 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 encontrará, la formativa se realiza durante el proceso [57].
- 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

2. ESTADO DEL ARTE

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 [63].

- **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 [80].
- **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 de si un criterio específico se ha cumplido o no [36].
- **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 esfuerzo y mejorar la motivación para aprender [56].

2.3.2.3 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:

- *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 [88]. 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 genérica.
- *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 genéricas.
- *Indicadores de trabajo en VLE*: indicadores del proceso de desarrollo del trabajo realizado por los estudiantes (no del trabajo final en sí). Encontramos estas técnicas en trabajos en los que tras haber trabajado los estudiantes en diferentes actividades del VLE son evaluados del desempeño en diferentes competencias a partir de los registros de interacción de estos estudiantes con el propio VLE.
- *Actividades sin determinar*: utilizaremos 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 VLE como actividades presenciales. No se

2. ESTADO DEL ARTE

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.

2.3.2.4 Instrumentos de evaluación

Los instrumentos de evaluación son las herramientas que se han utilizado para llevar a cabo las técnicas de evaluación descritas en el apartado anterior. En el siguiente listado podrá ver los instrumentos de evaluación utilizados en los trabajos seleccionados 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 [95].
- La *Entrevista* es una prueba oral en la que el estudiante deberá 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 [100] 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, ésta tiene lugar de una forma menos profunda e impersonal, que el “cara a cara” de la entrevista [71].
- La *herramienta de seguimiento* es una herramienta para monitorizar el trabajo del estudiante a lo largo del semestre. En el trabajo presentado en [60] el profesor 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 [102]. Uno de esos dominios es el laboral, sobre todo las entrevistas de trabajo. Es común la necesidad del empresario por conocer la aptitud o no 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 [34]. 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 ¹. 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 han extraído unas características comunes a la tipología de los trabajos.

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

¹<http://cs.gmu.edu/gaia/SeriousGames/index.html>

2. ESTADO DEL ARTE

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

Tabla 2.6: Distribución de publicaciones por tratamiento del problema

se propone en la introducción de este capítulo, es decir, aprovechando los registros de interacción de los estudiantes con el 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 si son los estudiantes los que se encargan de esa evaluación, mediante evaluaciones realizadas directamente por el profesor o mediante ambos enfoques. En concreto, de los 13 trabajos seleccionados que aplican evaluación entre iguales o autoevaluación, 4 aplican también evaluación del profesor. En la tabla 2.6 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 utilizará 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 trabajos típicos de un tema de investigación de cierta madurez como los de opinión (*Opinion papers*) y los filosóficos (*Philosophical papers*) casi no hay.

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 profesor (*teacher assessment*) y autoevaluaciones o evaluaciones entre compañeros (*peer and self-assessment*), mientras

que la segunda y la tercera se dan con más frecuencia con evaluaciones del profesor. Cabe destacar que en esta figura si hay un trabajo que utiliza dos tipos de contribución, por ejemplo, evaluaciones del profesor y evaluaciones entre iguales o autoevaluación, se contará una vez para cada tipo.

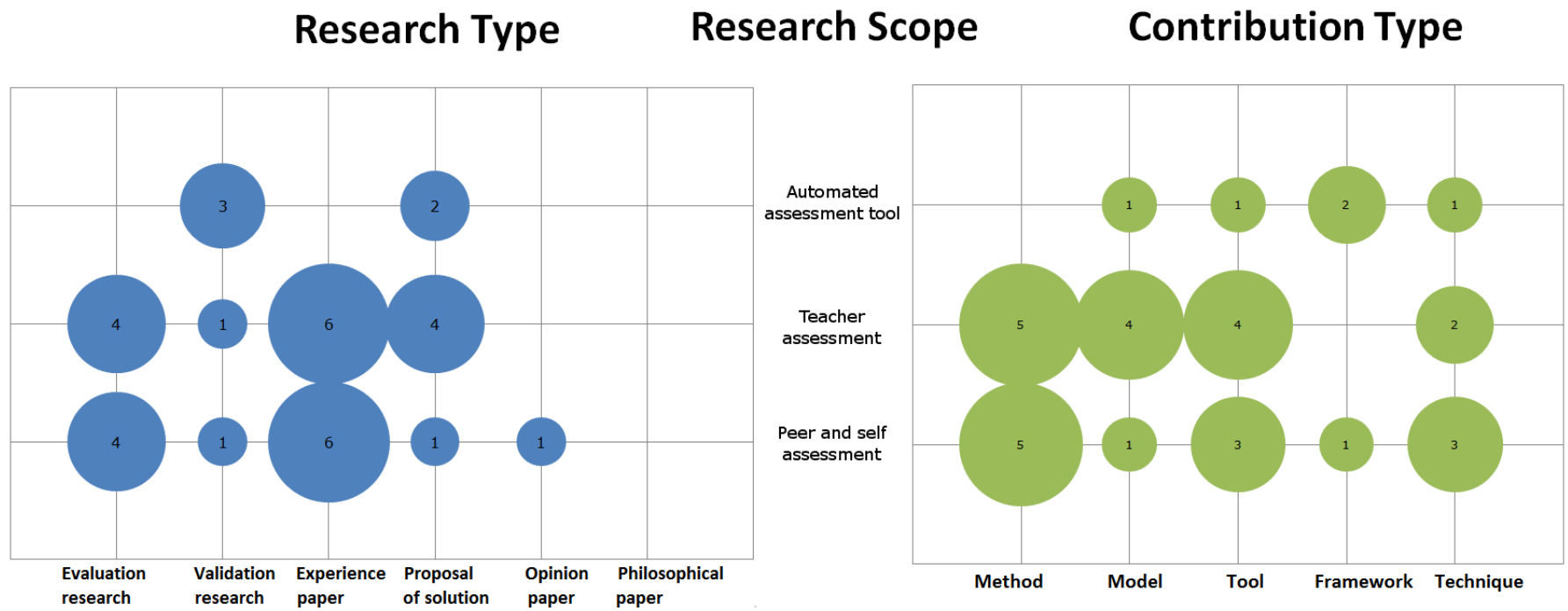


Figura 2.2: Ámbito de trabajos distribuidos según tipo de investigación y según tipo de contribución.

2.3.4 Esquema de clasificación

Todos los artículos seleccionados tienen en común dos cosas: 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 2.13.

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 profesor.
- Evaluación semiautomática: se utiliza una herramienta informática que evalúa de manera automática a los estudiantes. Puede ser que esta herramienta requiera una intervención inicial del profesor para introducir datos o configurar la propia herramienta, pero las evaluaciones se generan automáticamente.

2.3.4.1 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 ahorrar tiempo del profesorado. A menudo este tipo de evaluación se acompaña de algún tipo de rúbrica [64].

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

2. ESTADO DEL ARTE

mediante cuestionarios que completan los propios estudiantes o el profesor. Los profesores confirman que el uso de herramientas web mejora la participación en la comunidad favoreciendo la colaboración y la construcción de conocimiento compartido [93]. 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[6].

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 [66]. En [61] los profesores 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 otros trabajos los estudiantes realizan una experiencia de *aprendizaje basado en equipos* (TBL, del inglés, Team Based Learning) utilizando algún tipo de herramienta colaborativa. En [41] 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 [20] 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

autoevaluados o evaluados por sus compañeros en la competencia o competencias que debería desempeñar en la videoconferencia [67].

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 2.7, el *trabajo en equipo* (6 trabajos) y la *comunicación* (5 trabajos) son las competencias más evaluadas mediante evaluación entre iguales y autoevaluación. El típico caso que encontramos en este grupo es ese en el que los estudiantes trabajan en grupo y evalúan su desempeño y el de sus compañeros en el *trabajo en equipo* mediante una rúbrica [41]. 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 2.8 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 una vez terminado el trabajo que debían realizar en el VLE mediante rúbricas y cuestionarios.

Análisis

Aunque la autoevaluación y evaluación entre iguales son enfoques que quitan trabajo al docente, no todo son ventajas. En algunos trabajos anteriores, este enfoque a menudo sólo se utiliza de manera complementaria a algún otro tipo de evaluación [61, 91]. Además, se puede dar el caso que la autoevaluación no se ajuste del todo a la realidad del desempeño del estudiante. Por ejemplo, en [19] hay notables

2. ESTADO DEL ARTE

COMPETENCIA	CANT.	TRABAJOS
Análisis	2	[19, 61]
Aprendizaje permanente	1	[73]
Comunicación	5	[19, 66, 67, 73, 89]
Creatividad	1	[78]
Cultural	1	[73]
Emprendimiento	2	[1, 20]
Gestión de proyectos	1	[66]
Habilidades interpersonales	1	[66]
Investigación	1	[73]
Liderazgo	1	[66]
Pensamiento crítico	1	[6]
Planificación y gestión del tiempo	1	[66]
Resolución de problemas	1	[73]
Responsabilidad	2	[19, 89]
Segundo idioma	3	[67, 86, 91]
TIC	3	[61, 67, 73]
Trabajo autónomo	1	[61]
Trabajo en equipo	6	[19, 41, 61, 66, 73, 78, 89]

Tabla 2.7: Competencias genéricas evaluadas mediante autoevaluación y evaluación entre iguales

MÉTODOS DE EVALUACIÓN	TÉCNICAS DE EVALUACIÓN	INSTRUMENTOS DE EVALUACIÓN	TRABAJO
Auténtica	Actividades sin determinar	-	[86]
	Observación sistemática	Rúbrica	[73]
Formativa	Pruebas escritas	Cuestionario	[1]
	Observación sistemática	Rúbrica	[6, 78]
Sumativa	Pruebas orales		[19, 61]
	Pruebas escritas	Cuestionario	[67]
			[20, 41, 66, 89, 91]

Tabla 2.8: Instrumentos de evaluación y métodos correspondientes a los trabajos de la autoevaluación y evaluación entre iguales

diferencias entre las calificaciones que se auto-asignan los estudiantes en algunas competencias y las calificaciones que le asignaron los profesores 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 profesor. En esta experiencia se evaluaron cuatro competencias genéricas: *capacidad de análisis*, *habilidades de escritura*, *responsabilidad* y *capacidad de trabajo en equipo*. Cabe destacar discrepancias 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%.

2.3.4.2 Evaluación del profesor

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

2. ESTADO DEL ARTE

rúbricas para evaluar el trabajo de sus estudiantes en las herramientas del VLE 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 VLE [93], metodologías PBL [60] o experiencias basadas en videoconferencias [100].

En [103] se implementa un itinerario de aprendizaje incluyendo la evaluación de competencias genéricas en dicho itinerario. 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 las fórmulas necesarias para evaluar cada bloque del curso a partir de sus actividad, y cómo se combinan para evaluar los objetivos del curso y las competencias genéricas.

Competencias evaluadas

Las competencia genérica más evaluada mediante la evaluación del profesor es la *comunicación* oral y escrita (6 trabajos). En estos trabajos los profesores 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 2.9 puede ver la relación de competencias evaluadas y los trabajos.

Métodos, Técnicas e instrumentos de evaluación

Dentro de la evaluación del profesor 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 2.10.

2.3 Resultados

COMPETENCIA	CANT.	TRABAJO
Análisis	1	[7]
Aprendizaje permanente	1	[81]
Comunicación	6	[16, 60, 65, 81, 87, 103]
Emprendimiento	2	[81, 100]
Pensamiento crítico	2	[7, 60]
Planificación y gestión del tiempo	1	[60]
Resolución de problemas	5	[7, 16, 65, 87, 99]
Trabajo autónomo	1	[61]
Trabajo en equipo	5	[16, 60, 65, 81, 87]

Tabla 2.9: Competencias evaluadas directamente por el profesor

MÉTODOS DE EVALUACIÓN	TÉCNICAS DE EVALUACIÓN	INSTRUMENTOS DE EVALUACIÓN	TRABAJO
Auténtica	Actividades sin determinar	-	[93]
Con referencia al criterio			[90]
Formativa	Pruebas orales	Entrevista	[100]
	Observación sistemática	Herramienta de seguimiento	[16, 60]
		Rúbrica	[87]
Sumativa	Actividades sin determinar	-	[7, 65, 81]
			[103]
	Pruebas escritas	Cuestionario	[99]

Tabla 2.10: Instrumentos de evaluación y métodos correspondientes a los trabajos en los que la evaluación es realizada por los profesores

2. ESTADO DEL ARTE

Análisis

La escalabilidad es el problema más mencionado por los autores en los trabajos recopilados. En [90] 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 profesores evaluaban las competencias de los estudiantes manualmente y después aplicaban *Hiperion*. La principal desventaja de la herramienta es el tiempo que el profesor 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 [60]. 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 profesor es un poco mayor al habitual. Lo mismo ocurre en [16], trabajo en el que los profesores 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.

2.3.4.3 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 profesor para introducir datos o configurar la herramienta.

En [4] 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.

Encontramos trabajos en los que se emplea una metodología de aprendizaje basado en juegos (GBL, del inglés, Game-Based Learning) [14, 50]. En ellos se utilizan juegos digitales con el fin de apoyar y mejorar la enseñanza, el aprendizaje y/o la evaluación. Se puede establecer como una práctica donde los jugadores, a

medida que van avanzado en las dinámicas del juego, deben evidenciar unas habilidades, conocimientos y competencias que muestren el alcance de los objetivos de aprendizaje [21].

Y finalmente encontramos trabajos que se basan en técnicas de *learning analytics* [42, 82], traducido al español como análisis del aprendizaje. El *learning analytics* es definido por la *Society for Learning Analytics* como la medición, recopilación, análisis y presentación de datos sobre los estudiantes, sus contextos y las interacciones que allí se generan, con el fin de comprender el proceso de aprendizaje que se está desarrollando y optimizar los entornos en los que se produce [92].

Competencias evaluadas

En la tabla 2.11 se muestran las competencias evaluadas y los trabajos en que se evalúan. Aunque son pocos trabajos los que encontramos en este grupo, cabe destacar que vuelve a ser la competencia de la *comunicación* la más evaluada.

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 2.12.

Análisis

La creación de los tests de personalidad no está al alcance de todos los profesores. Como hemos visto antes en el trabajo presentado en [4], fue necesario llevar a cabo un modelo Delphi para la definición de los tests.

En [50] 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 [14] 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

2. ESTADO DEL ARTE

COMPETENCIA	CANT.	TRABAJO
Análisis	1	[4]
Aprendizaje permanente	1	[4]
Comunicación	3	[4, 14, 82]
Creatividad	1	[4]
Emprendimiento	1	[50]
Gestión de proyectos	1	[4]
Habilidades interpersonales	2	[4, 82]
Investigación	1	[4]
Liderazgo	1	[4]
Pensamiento crítico	1	[4]
Planificación y gestión del tiempo	1	[4]
Resolución de problemas	1	[50]
Responsabilidad	1	[4]
Toma de decisiones	1	[4]
Trabajo autónomo	1	[4]
Trabajo en equipo	2	[4, 42]

Tabla 2.11: Competencias evaluadas de forma semiautomática

MÉTODOS DE EVALUACIÓN	TÉCNICAS DE EVALUACIÓN	INSTRUMENTOS DE EVALUACIÓN	TRABAJO
Auténtico	Indicadores basados en logros	Juegos serios	[14]
Formativo			Indicadores de trabajo en VLE
	[42, 82]		
Sumativo	Pruebas escritas	Test automático	[4]

Tabla 2.12: Instrumentos de evaluación y métodos correspondientes a los trabajos de evaluación semiautomática

utilizarse con un propósito específico, generalmente relacionado con competencias específicas. Hay muchos trabajos sobre juegos serios en la literatura pero muy poco sobre competencias genéricas.

En el primer trabajo en el que se utilizan los registros de actividad de los entornos de aprendizaje se 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 [82]. 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, es decir, número de acciones que se han hecho a tiempo con respecto a la planificación fijada en una hoja de cálculo de Google.
- *Comunicación interpersonal*. Indicador: Escuchar a los demás, es decir, número de veces que los estudiantes accedieron a las discusiones del foro de Moodle.
- *Comunicación interpersonal*. Indicador: Expresar sus ideas, es decir, intervenciones en el foro de Moodle y comentarios en el documento Google.
- *Habilidades de escritura*. Indicador: Expresar sus ideas, es decir, número de veces que el 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, es decir, número de total de palabras que cada estudiante escribió dividido por el número de veces que intervino.
- *Trabajo en equipo*. Indicador: Participación activa, es decir, tiempo total que duran las discusiones divididos por el número de sesiones.
- *Pensamiento analítico*. Indicador: Respaldo de ideas de otros, es decir, número total de post 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, es decir, 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*), un método que integra herramientas que están presentes en diferentes entornos de aprendizaje virtual y que facilita el registro de la interacción de los estudiantes y por tanto un acceso más sencillo a las evidencias del trabajo en equipo [42]. El registro de esta interacción en el foro era una tarea tediosa para ser realizada a mano, por lo que implementaron el *LA system*. *LA system* se implementa como un servicio web en Moodle, de forma que la información sea accesible a través de internet con mensajes basados en XML. Esta información después es consumida por un cliente que proporciona una visión apropiada de los datos. Los indicadores utilizados son:

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

Ambos trabajos nos 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 curso virtual 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 profesor de dicho curso virtual. Sería deseable la opción del profesor de poder diseñar sus propias evaluaciones a partir de los indicadores que proporciona el entorno virtual.

REF	TÍTULO	TIPO DE INVESTIGACIÓN	TIPO DE CONTRIBUCIÓN	ÁMBITO DE LA INVESTIGACIÓN
[103]	A Fine-Grained Outcome-Based Learning Path Model	proposal of solution	model	Teacher assessment
[82]	A web platform for the assessment of competences in Mobile Learning Contexts	validation research	framework	Automated assessment tool
[65]	Acquired Skills With The Implementation Of New Evaluation Methods At University Rey Juan Carlos	experience paper	model	Teacher assessment
[60]	Active learning through problem based learning methodology in engineering education	experience paper	method	Teacher assessment
[16]	Adapting teaching and assessment strategies to enhance competence-based learning in the framework of the european convergence process	proposal of solution	method	Teacher assessment
[7]	Appraisal of Course Learning Outcomes using Rasch Measurement: A Case Study in Information Technology Education	proposal of solution	model	Teacher assessment
[91]	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
[99]	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
[94]	Criteria and standards of generic competences at bachelor degree level: A review study	evaluation research	method	Review study

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

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

Tabla 2.13: Distribución de publicaciones por tratamiento del problema

2. ESTADO DEL ARTE

2.4 Respuestas a las preguntas de investigación

En base al estudio mostrado 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?

En esta selección de trabajos se han encontrado evaluaciones para todas las competencias que fueron definidas en la tabla 2.5. Aunque las que más se han evaluado 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 2.14.

Estos datos se pueden contrastar con los de la revisión de la literatura mostrada en [94], donde se analizan las competencias genéricas más frecuentemente evaluadas. En este caso repite 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 este caso 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 VLE favorecen los procesos de cooperación [51].

Q2. ¿Qué métodos se utilizan para evaluar competencias genéricas mediante el uso de entornos virtuales?

Se han encontrado dos grupos de trabajo, por un lado, aquellos en los que el entorno virtual asiste al usuario en la evaluación de competencias genéricas, y por el otro, aquellos en los que el entorno virtual 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 profesor.

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

2.4 Respuestas a las preguntas de investigación

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

Tabla 2.14: Número de trabajos que evalúan cada competencia genérica

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 profesor (5 trabajos) con respecto a trabajos que evalúan los estudiantes (1 trabajo).

La herramienta que más se utiliza tanto en la evaluación del profesor 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 profesor se encarga de la evaluación, la carga de trabajo de éste aumenta [60]. 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 realmente merecen [19].

Dentro de la evaluación semiautomática se seleccionaron 5 trabajos, entre los que se encuentran algunos basados en juegos serios [14, 34] y otros basados en el análisis de los procesos de aprendizaje [42, 82].

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

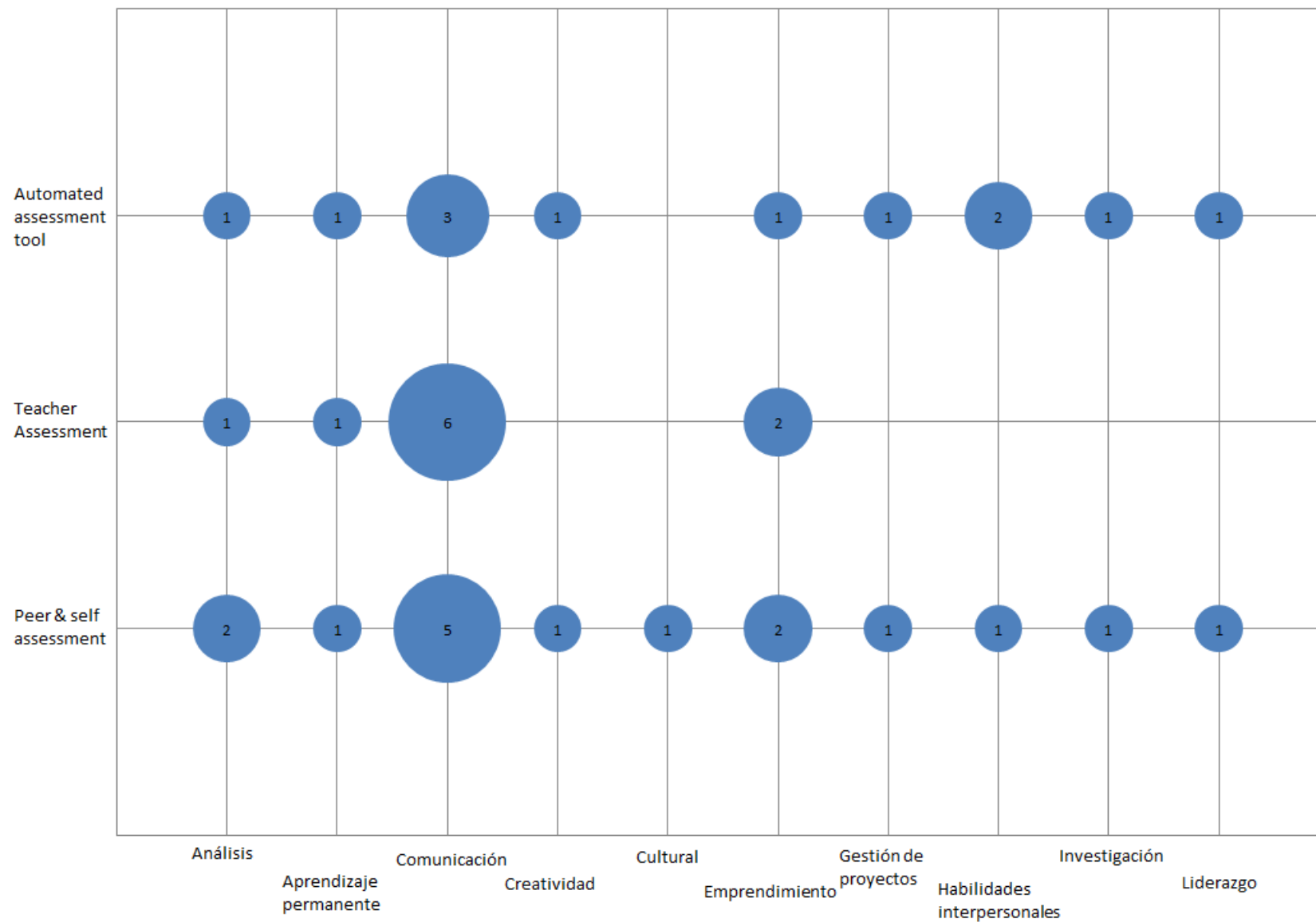


Figura 2.3: Competencias genéricas evaluadas con cada método (1 de 2).

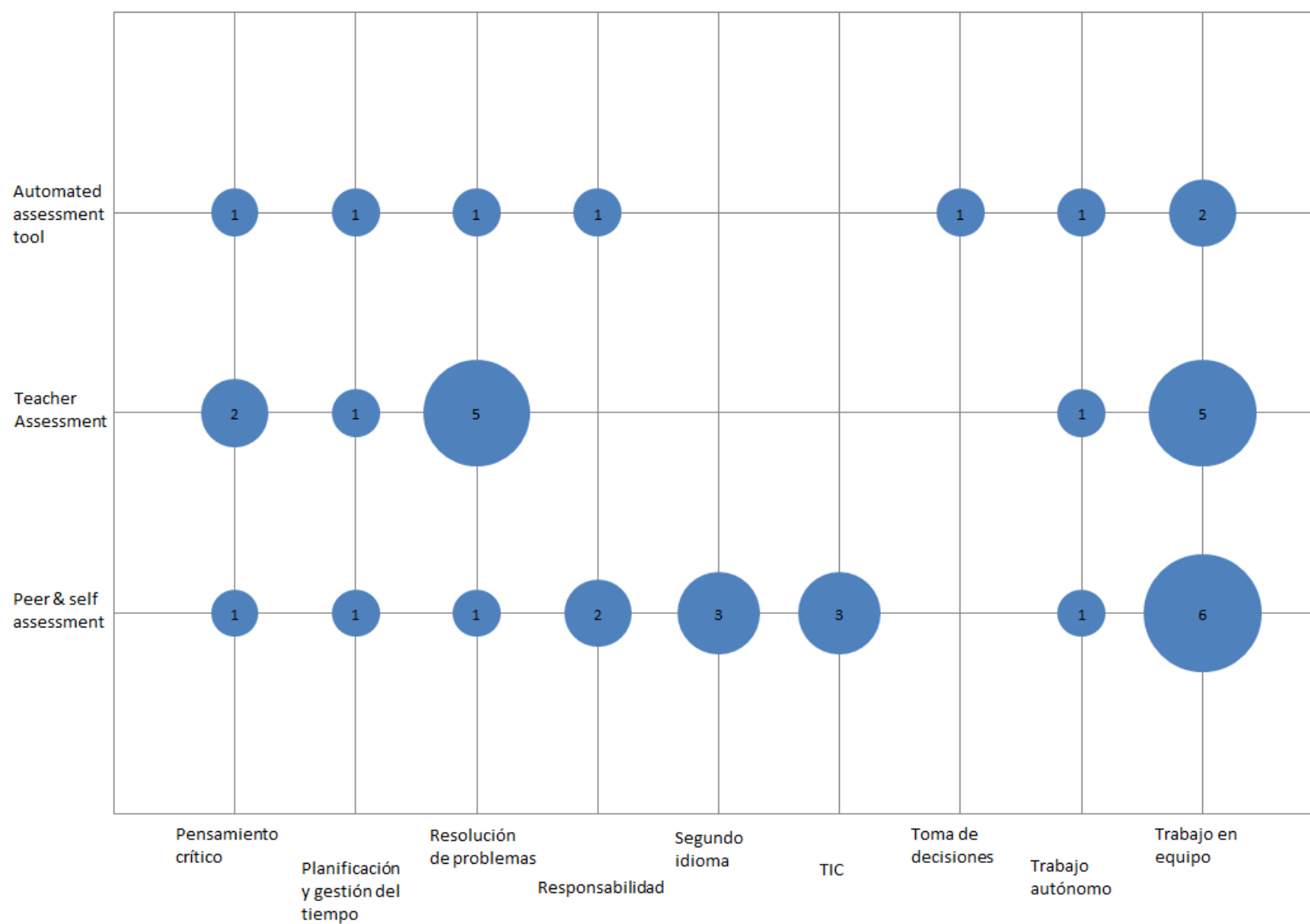


Figura 2.4: Competencias genéricas evaluadas con cada método (2 de 2).

2.4 Respuestas a las preguntas de investigación

Competencias	Métodos	Técnicas	Herramientas
	Formativo	Indicadores de trabajo en VLE	Herramienta para el análisis de los registros de aprendizaje
<i>Análisis</i>	[82]		
<i>Comunicación</i>			
<i>Habilidades interpersonales</i>			
<i>Planificación</i>			
<i>Trabajo en equipo</i>	[42]		

Tabla 2.15: Competencias evaluadas, métodos aplicados y técnicas seguidas para evaluar competencias a partir de los registros de actividad de los entornos de aprendizaje

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

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 VLE utilizando herramientas para el análisis de los registros de aprendizaje [42, 82].

En [82] 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 [42] se utilizan procesos de minería de datos (*data mining*) mediante el desarrollo de un servicio web en Moodle que devolviera información almacenada en los registros en la plataforma de aprendizaje.

A continuación se muestra un resumen para estos trabajos en el que se muestra las técnicas empleadas, los métodos aplicados y las competencias evaluadas 2.15.

2.5 Conclusiones

Las competencias genéricas son hoy en día una pieza fundamental en las planificaciones de las asignaturas 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 estos trabajos y proyectos, es necesaria 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 que generar y evaluar nuevas actividades para evaluar las competencias genéricas. Además, este problema de escalabilidad se acrecienta aún más en el contexto de los entornos de aprendizaje virtuales, donde los cursos en ocasiones contienen un número muy elevado de estudiantes (por ejemplo los cursos de tipo MOOC (*Massive Online Open Courses*)).

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 profesor quién 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 profesor combinando o sustituyendo la evaluación del profesor con evaluaciones entre iguales o autoevaluaciones. Esto esquiva 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 volvemos a encontrarnos 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

carácter o mecanismos psicológicos de un individuo. El inconveniente es que realizar un cuestionario de este tipo no está al alcance de cualquiera, sino que debe ser realizado por psicólogos. Además, son éstos los que deben diagnosticar o valorar los resultados. Y aunque se trate de automatizar, en ocasiones estos tests suelen contar con preguntas cerradas y abiertas, y en el caso de esta última, volvemos a encontrarnos con problemas de escalabilidad.

Por último nos hemos encontrado con un conjunto de trabajos que automatizan el proceso de evaluación de competencias genéricas. Por un lado nos encontramos con 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 genéricas. 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 profesor, 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. En este conjunto de trabajos se 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 es válido para un profesor podría no serlo para otro. Si tuviéramos un método que nos permitiera diseñar evaluaciones podríamos probar diferentes fórmulas para obtener diferentes indicadores del desempeño de los estudiantes en el entorno virtual hasta encontrar aquellos que se adapten a lo que cada profesor considere válido para la evaluación de competencias.

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

Winston Churchill

CAPÍTULO

3

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 técnicas informáticas, así como de cuáles son esos métodos y si se están usando para este fin los registros de actividad de los entornos virtuales. A raíz de esta revisión han aflorado varios problemas para cada tipo de evaluación encontrado que van a ser resumidos a continuación.

3.1 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 profesor (evaluación del profesor).

3. RESUMEN DE PROBLEMAS ENCONTRADOS

3.1.1 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. 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 profesor.

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 realidad, habiendo diferencias notables entre las calificaciones que asignan diferentes estudiantes o el profesor a un mismo trabajo. Este tipo de evaluaciones suelen ir acompañados de rúbricas para guiar el proceso de evaluación [64]. Sin embargo, no todos 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 en la materia que pueden tener los estudiantes o una interpretación diferente que estos realicen de los criterios de evaluación, las diferencias entre la calificación que el profesor daría a un trabajo y la que darían sus estudiantes podrían ser importantes [19].

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. Es cierto que la autoevaluación o evaluación entre iguales son estrategias que ahorran trabajo al profesor, ya que son los estudiantes los que se encargan de parte de dicha evaluación. Sin embargo, en numerosos trabajos, tras la evaluación realizada por los estudiantes los profesores han de revisar las evaluaciones de sus estudiantes, por lo que al final no sólo no se ahorra tiempo, sino que puede que tengan que realizar un mayor número de evaluaciones si cada estudiante ha corregido más de un trabajo [61, 91]. Más sentido tiene aún esta revisión del profesor si

tenemos en cuenta problemas como el descrito anteriormente, en los que las calificaciones de los estudiantes puede que no se ajusten a la realidad.

3.1.2 Evaluación del profesor

En esta sección se recogen problemas asociados a los trabajos en los que la evaluación la realiza directamente el profesor 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 [60, 90] e incluso uno de ellos descarta repetir la experiencia en los siguientes cursos debido al sobreesfuerzo que le supuso [16]. Podemos deducir, por tanto, que si en ocasiones los profesores apenas tienen tiempo de lograr los objetivos curriculares en cuanto a 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 evaluar dichas tareas.

3.2 Evaluación semiautomática

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

3. RESUMEN DE PROBLEMAS ENCONTRADOS

Propósito específico

Los recursos con los que cuentan los profesores 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 [14, 50]. Además, el desarrollo de un juego para evaluar las competencias concretas que a un profesor le interesa no es una tarea que este al alcance de cualquier profesor. 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 con un perfil relacionado con la psicología, algo que obviamente no tienen todos los profesores. 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 profesores [4].

Si hablamos de los juegos serios nos ocurre lo mismo, un profesor 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 y que este al alcance de cualquier profesor.

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 [42, 82]. El problema es que los indicadores son fijos, y lo que para un profesor 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 profesor 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.

3.2 Evaluación semiautomática

Es decir, lo que sería deseable es que hubiera un método para que el profesor pueda diseñar sus propias evaluaciones en base a estos indicadores.

*Historical methodology, as I see it,
is a product of common sense ap-
plied to circumstances.*

Samuel E. Morison

CAPÍTULO

4

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 describe el método, sus características, los requisitos que el método ha de satisfacer y se introduce el tipo de herramienta que se empleará para su implementación.

4.1 Introducción

Se podría decir que hoy en día los entornos virtuales constituyen una pieza fundamental en cualquier contexto en el que se impartan cursos. Mientras que en los cursos virtuales, los VLEs son el único entorno de trabajo posible, en los cursos presenciales o mixtos, tanto los VLEs como otras herramientas virtuales 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 indicadores de la actividad generada por los estudiantes en los entornos de aprendizaje virtual. Los profesores podrán diseñar evaluaciones utilizando estos indicadores y podrán utilizar estas evaluaciones como evidencias del desempeño de competencias genéricas. Si llegado el caso, el profesor considera que las evidencias no terminan de reflejar bien el desempeño o no de las competencias, el método DBA permitirá al profesor refinar los diseños hasta llegar a unas evidencias que sí le sean válidas al profesor, o a descartarlas si no llegan a serlo. Además, los diseños podrán ser utilizados y modificados por otros profesores que busquen adaptarlo a su contexto local o a las competencias genéricas que ellos quieren medir.

En el VLE por ejemplo, los estudiantes pasan a diario por sus páginas. Por un lado, habrá estudiantes que accedan al VLE 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 de los estudiantes quedan almacenadas en el registro de actividad de los entornos virtuales, 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 indicadores procedentes de estos entornos virtuales.

4.2 Método: design-based assessment (DBA)

4.2.1 Contexto

Los estudiantes comienzan a dejar constancia de su actividad en los entornos virtuales desde el momento en que acceden al entorno. El registro del sistema lo almacena todo, tanto la participación activa del estudiante, cuando éste interactúa con el entorno virtual, como la participación pasiva, cuando el estudiante simplemente accede y navega de una página a otra.

Los programas de las asignaturas incluyen las competencias genéricas de las que los estudiantes deben ser evaluados. Los profesores pueden plantear actividades en los entornos virtuales con la intención no sólo de evaluar ciertas habilidades de

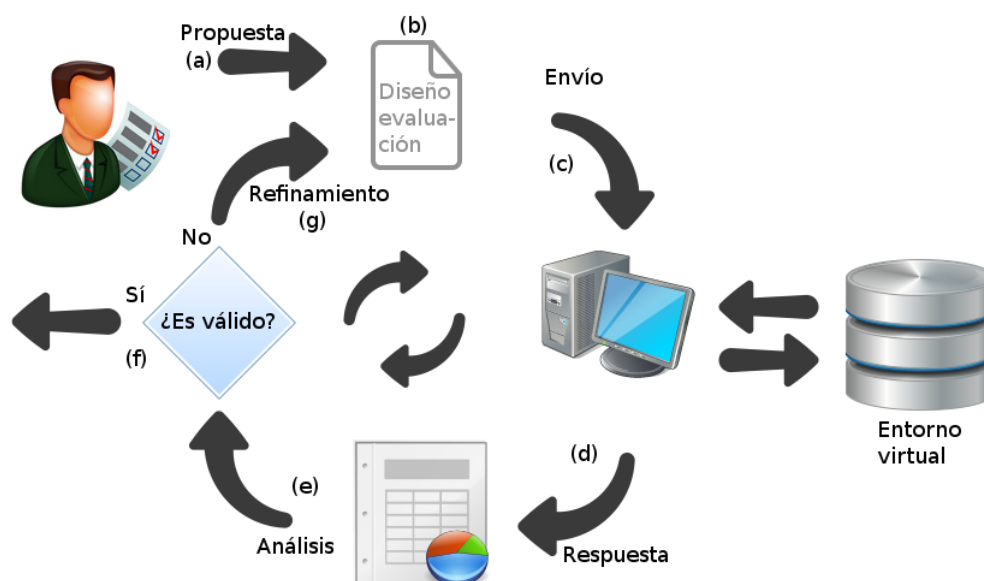


Figura 4.1: Diagrama del ciclo de contraste de hipótesis

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.

Para evaluar una competencia genérica dada, el profesor podrá 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 profesor formula una hipótesis para la utilización de algún tipo de información de la actividad de los estudiantes en el entorno

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

virtual 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 profesor en el VLE con anterioridad a la fecha fijada para las mismas.

2. *Diseño y formulación de evaluación*: el profesor diseñará 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, 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 enviará 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. *Validación de resultados*: la herramienta pondrá a disposición del profesor los indicadores requeridos (d). El profesor los analizará (e) y evaluará si son válidos para el propósito que fueron diseñados (f), si necesitan ser refinados o si hay que descartarlos. 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

En este ejemplo, el profesor pretende obtener indicadores del uso del foro del entorno virtual para evaluar la competencia genérica de las habilidades interpersonales. Durante el curso, el profesor planteó una actividad de debate mediante el foro de la asignatura que tendría lugar durante una semana de las que conforman

4.2 Método: design-based assessment (DBA)

el curso. El profesor aplicará dos iteraciones del ciclo de contraste de hipótesis del método DBA.

Iteración I

I.1 Hipótesis inicial. El profesor 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.*

I.2 Diseño y formulación de la evaluación. Para diseñar la evaluación, el profesor 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 4.1, escrita en el lenguaje SASQL ¹ y que proporcionará datos sobre la participación de los estudiantes en el foro entre dos fechas.

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

Consulta 4.1: Participación en el foro en un periodo concreto de tiempo

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

I.4 Validación de resultados. El software devuelve al profesor los indicadores solicitados mediante el listado 5.2. El profesor 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 profesor considera pobre esta primera aproximación y decide redefinir la hipótesis inicial.

¹<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

Iteración II

II.1 Hipótesis inicial. El profesor 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.*

II.2 Diseño y formulación de la evaluación. El profesor diseña la nueva evaluación mediante la consulta 4.2.

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

Consulta 4.2: Interacción en el foro en un periodo de tiempo

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

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 vemos que sólo dos de los estudiantes cumplen la segunda hipótesis (*student2* y *student4*), y bajo esta nueva hipótesis, ellos serian los dos únicos estudiantes que han mostrado un buen desempeño de la competencia genéricas de las habilidades interpersonales.

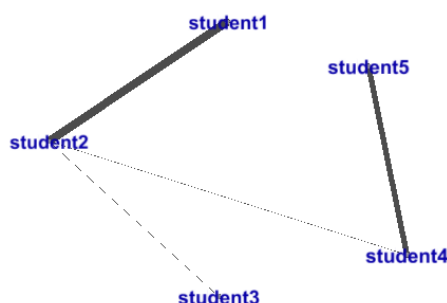


Figura 4.2: Interacción en el foro en un periodo de tiempo.

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 2.3.2.2 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 profesor.

Mientras que con respecto a la clasificación de técnicas de evaluación mostrada en la sección 2.3.2.3, la técnica empleada en este método es la de obtención de *indicadores del trabajo en actividades de aprendizaje*.

4.3.1 Características

En el apartado 1.1 se indican, partiendo del análisis realizado por Terry Andersen y Julie Shattuck [3], las características que un estudio DBR de 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.

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

- *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 [3].
- *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 mediante el método DBA de sus habilidades interpersonales. Esa actividad podría formar parte de un conjunto de actividades de trabajo en equipo configuradas por el profesor para realizar una evaluación auténtica. El profesor 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 profesor o hasta que decida descartarlos por no ser válidos para su evaluación.
- *Asociación colaborativa entre investigadores y profesores:* 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 juntos para que se cumplan las expectativas. Esto es especialmente apropiado en el campo de los DSLs, herramientas informáticas diseñadas para facilitar el desarrollo de software en un contexto específico [58]. Bajo este enfoque se desarrollará el método DBA, colaborando profesores e investigadores en su desarrollo, teniendo así los usuarios finales y expertos en el dominio (los profesores) una participación activa y directa en la definición de una herramienta que implemente el método, es decir, un DSL.

4.3 Características, requisitos y herramienta

- *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 profesores, y a partir de su reflexión, modificados para ser adaptados a nuevos contextos. De esta forma, los principios y diseños creados por profesores e investigadores no mueren con una práctica, sino que son puntos de partidas para nuevas teorías y experimentos. Por ejemplo, un profesor que trabajase con wikis en sus clases podría utilizar como indicador de un buen desempeño de la competencia de trabajo en equipo el número de contribuciones aportado por cada estudiante a la página de su equipo de trabajo. Un segundo profesor 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 profesor podría mejorar esta evaluación añadiendo un indicador cualitativo de la información aportada al wiki por cada miembro del equipo. Es decir, se parte de un patrón o diseño inicial que los profesores 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 profesores mientras que, por lo general, la práctica de la investigación-acción es llevada a cabo únicamente por el profesor que realiza el experimento. Ambos enfoques son muy parecidos y en ocasiones cuesta diferenciarlos. Se podría decir que la investigación-acción carece de ese carácter reflexivo que sí tiene el DBR [25]. 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 profesor 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

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

ocurrió en el ejemplo anterior de wikis, donde cada profesor, investigador o equipo de investigadores y/o profesores 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 las sesiones prácticas y del análisis de resultados, los investigadores y profesores 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

El método debe cumplir una serie de requisitos que parten de los inconvenientes encontrados en la revisión de la literatura realizada en el capítulo 2 y que han sido resumidos en el capítulo 3. A continuación se describe cada uno de estos requisitos.

1. Indicadores objetivos

Los indicadores reflejan los datos obtenidos directamente del registro de las actividades de aprendizaje, por lo que serán objetivos per se. No ha lugar a consideraciones personales o interpretaciones inexactas de rúbricas cómo ocurría en la autoevaluación o evaluación entre iguales, dónde dos evaluaciones de un mismo trabajo realizadas por personas diferentes podrían tener calificaciones diferentes. En el caso de los indicadores obtenidos del registro de las actividades de aprendizaje, dos estudiantes que tienen los mismos datos en el registro tendrán en el mismo valor en el indicador, y ya será decisión del profesor la interpretación de este indicador en relación con otros indicadores.

2. Evaluación escalable

El método para la evaluación de competencias genéricas deberá ser escalable, sin que suponga al profesor un esfuerzo que éste no pueda abordar. El método se alineará con las actividades aprendizaje para que el profesor pueda consultar los

4.3 Características, requisitos y herramienta

indicadores del registro con una simple petición a la herramienta. La herramienta procesará la petición y devolverá la información en formatos que el profesor podrá visionar y exportar a otras herramientas.

3. Propósito general

El propósito del método es obtener indicadores del registro de las actividades de aprendizaje y que estos sean utilizados para evaluar competencias genéricas. Pero no están orientados a una competencia genérica concreta, sino que es el profesor quien diseña sus actividades en el entorno virtual y el que luego obtiene los indicadores para después utilizarlos en la evaluación de la competencia genérica que considera que los estudiantes han desempeñado en dicha tarea (y que queda reflejada en los indicadores). El profesor podría incluso utilizar los indicadores para evaluar competencias específicas si lo creyese oportuno. Pero en ningún caso este método tiene como propósito una competencia y actividad concreta como ocurría, por ejemplo, con algunos juegos serios recogidos en el estado del arte.

4. Accesibilidad

No deberá ser un requisito que el profesor 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 profesores puedan utilizarla sin requerirles conocimientos técnicos, 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

La herramienta proporciona al profesor la posibilidad de diseñar sus propias evaluaciones a partir de los indicadores. En el estado del arte nos encontramos con trabajos que obtenían sus evaluaciones a partir de los indicadores del VLE, pero éstos eran fijos. Es decir, cada competencia se evaluaba con un indicador dado. Pero podía ocurrir que el profesor no utilizase las actividades del VLE que proporcionaban dichos indicadores o que plantease las actividades con un enfoque diferente

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

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 profesor plantea actividades en el wiki de manera que cada estudiante trabaje individualmente en una página, podrá valorar otras competencias, pero el trabajo en equipo no. Con este método el profesor es quién diseña sus indicadores, y por tanto, sus evaluaciones a partir de los registros de las actividades de aprendizaje.

En resumen, podemos decir que el método que se propone para evaluar competencias genéricas a partir de los registros de interacción de las actividades de aprendizaje se pone a disposición del profesor en forma de una herramienta informática que se conecta a la actividad de aprendizaje utilizado en la asignatura. Mediante esta herramienta, los profesores pueden diseñar evaluaciones a partir de indicadores objetivos obtenidos del VLE y aplicarlos a las competencias genéricas para las que ellos consideren que les son válidos.

4.3.3 Herramienta

La herramienta que implemente el método debe cumplir una serie de requisitos que ayuden a alcanzar los mencionados en el punto anterior:

- *Cercana al dominio*: la herramienta debe servir al profesor o investigador para definir indicadores a partir de la actividad de los estudiantes en los entornos de aprendizaje, por tanto, debe ser una herramienta definida en términos del ámbito educativo.
- *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.
- *Usabilidad*: la herramienta debe ser usable por usuarios de perfil no informático, por lo que su utilización debe ser sencilla y entendible por este tipo de usuarios.

4.3 Características, requisitos y herramienta

A tenor de estos requisitos se decidió que la herramienta informática a implementar sería un *lenguaje específico de dominio* (DSL, del inglés, *domain-specific language*). 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 [45]. En el método DBA, el ejecutable sería el diseño y los expertos en el dominio los profesores.

*The logic of validation allows us
to move between the two limits of
dogmatism and skepticism.*

Paul Ricoeur

CAPÍTULO

5

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.

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 cuarto lugar, se presentan las contribuciones surgidas de esta tesis doctoral. 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 VLE (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)¹, 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)². SMW realizaba un análisis de las contribuciones de los estudiantes en el wiki meramente cuantitativo, pero que en su aplicación se topó con algunos inconvenientes que requerían una aproximación cualitativa.

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

5.2.1 Introducción

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 tradiciona-

¹<http://assessmediawiki.forja.rediris.es>

²<http://statmediawiki.forja.rediris.es/>

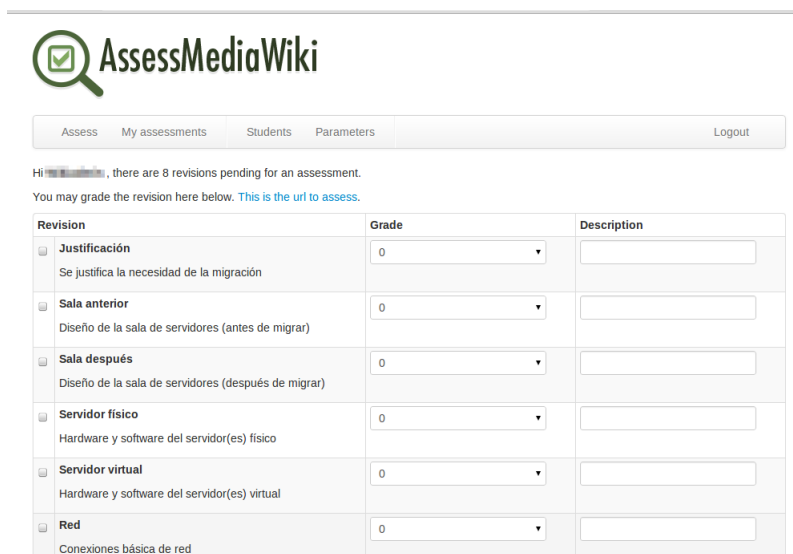
les [37]. Algunas de las ventajas sobre los medios tradicionales, ya sean en formato impreso o en documentos digitales, es que éstos no llevan un registro de ediciones, no permiten la colaboración distribuida y asíncrona y no pueden ser monitorizados por el docente mientras los estudiantes completan el trabajo.

Para evaluar el trabajo final de un grupo de estudiantes en una página del wiki nos bastaría con leer la última versión de dicha página, como hacíamos con los métodos tradicionales. Pero 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 [96]. 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 [74]. 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 imposible de abordar 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 [35]. 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 [76].

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 web application. At the top, there is a logo with a magnifying glass over a checkmark and the text "AssessMediaWiki". Below the logo is a navigation bar with links: "Assess", "My assessments", "Students", "Parameters", and a "Logout" button. A message states: "Hi [username], 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 representing a revision to be assessed.

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

Figura 5.1: Rúbrica de 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 (figura5.1).

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 vigilar 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% más significativa que aún no ha sido evaluada.

Al revisar sus evaluaciones, los estudiantes puede revisar las notas recibidas y sus justificaciones, así como ver a qué contribución en particular se refiere (figura 5.2). Si el estudiante no está de acuerdo con la calificación puede replicar utili-

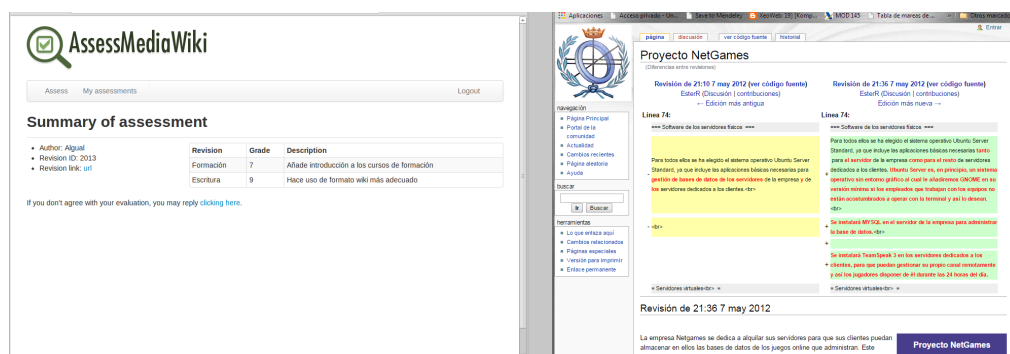


Figura 5.2: Ejemplo de retroalimentación formativa y la contribución de wiki evaluada

zando 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 la disputa y pondrá la nota definitiva.

5.2.2.1 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, una segunda fase de evaluación y una tercera fase de revisión del docente. 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.

Fase 1: 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

5. RESULTADOS DE LA INVESTIGACIÓN

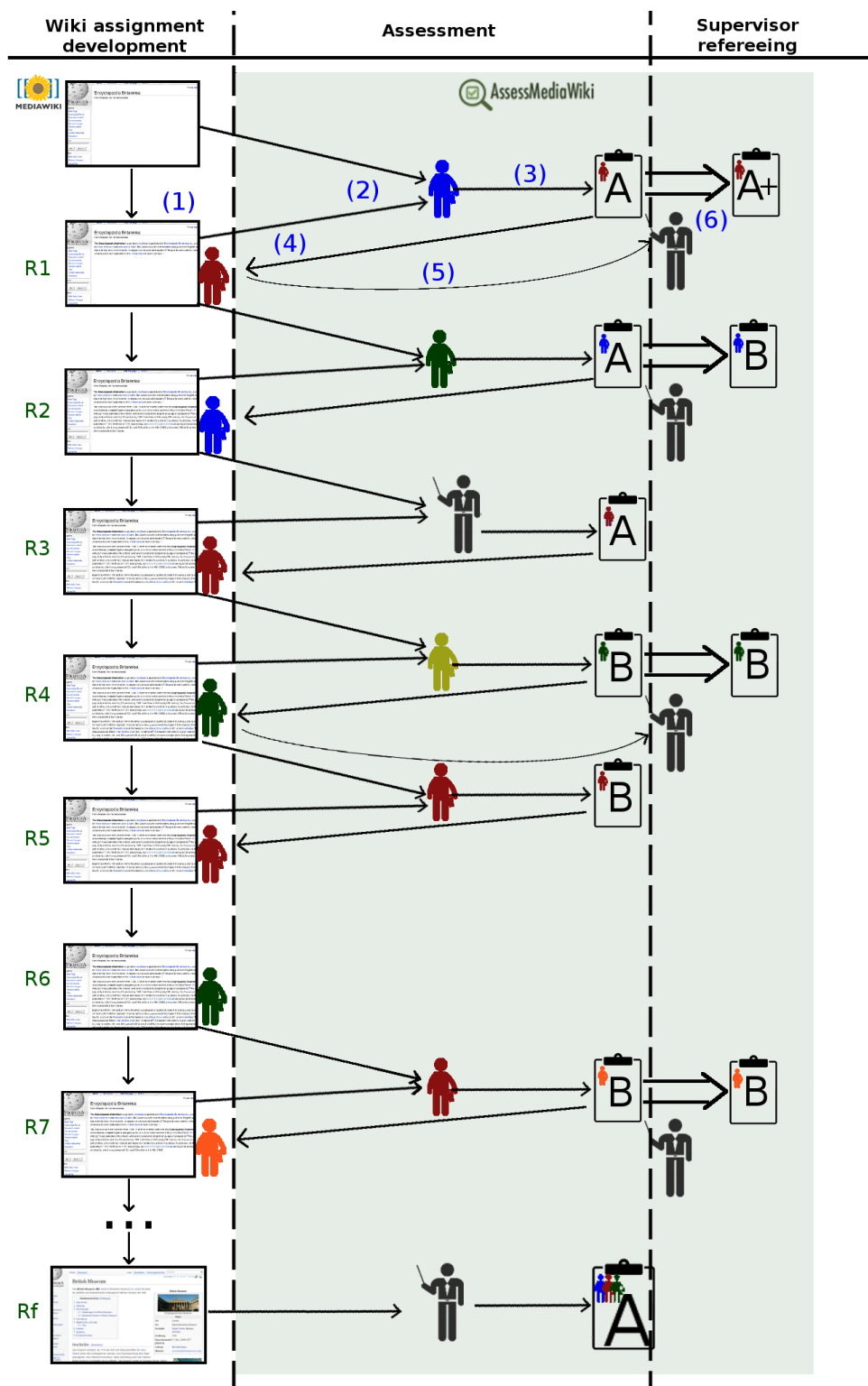


Figura 5.3: Ejemplo de flujo de trabajo para la evaluación cualitativa del wiki utilizando AMW

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 (Rf 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 o no.

Fase 2: Evaluación. Esta fase se muestra en la columna central y comprende las siguientes actividades:

- *Autoevaluación, evaluación entre iguales y evaluación del docente.* Las contribuciones a ser evaluadas se asignan a los estudiantes. Cada contribución es la diferencia entre dos revisiones consecutivas de una página del wiki. El estudiante encargado de evaluar dicha contribución se representa en el gráfico como un usuario coloreado que recibe dos flechas de las revisiones, una de la revisión anterior a 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 atómica 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 contribución 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 a la página del wiki por el estudiante rojo (I) al estudiante azul. El estudiante

5. RESULTADOS DE LA INVESTIGACIÓN

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 vemos como 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 estudiante 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*. Está claro que sería deseable que todas las contribuciones significativas fueran evaluadas, pero no es escalable.

- *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 injusta 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 necesaria 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 descartadas por la función de selección 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 son multidimensionales.

Un componente interesante de nuestro algoritmo es qué contribución 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). Por lo que cada evaluación adicional a la misma contribución provocará que otras contribuciones sean más pobremente evaluadas o que no lo sean.
- *¿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

5. RESULTADOS DE LA INVESTIGACIÓN

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 puede solicitar que 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 balancear el esfuerzo requerido y el detalle a exigir en las evaluaciones.

Fase 3: 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*. En la figura 5.4 puede verse la descripción del ciclo. Esta

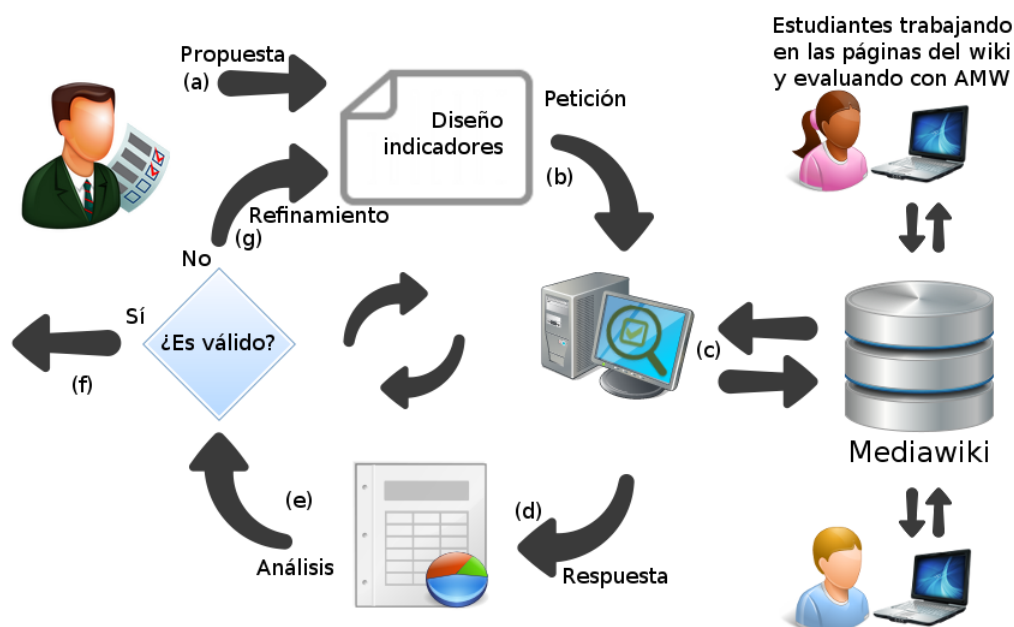


Figura 5.4: Ciclo de contraste de hipótesis para la evaluación en wikis

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.

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

5. RESULTADOS DE LA INVESTIGACIÓN

primer momento, este método proporciona indicadores y es el docente el que los utilizará para evaluar las competencias genéricas que considere oportunas.

Trabajo en equipo. El indicador considerado para el *trabajo en equipo* 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 *trabajo en equipo*.

Comunicación y aplicación del conocimiento. El indicador considerado para la comunicación y la aplicación del conocimiento 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.

Mantener la calidad del trabajo producido. El indicador considerado para el mantenimiento de la calidad del trabajo producido 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.

5.2 Antecedentes

COMPETENCIAS	INDICADORES CUALITATIVOS	INDICADORES CUANTITATIVOS
Trabajo en equipo	Ratio de miembros del equipo que trabajaron en un mismo criterio	Ratio de miembros del equipo que contribuyeron a una misma página del wiki en las páginas de su proyecto
Comunicación y aplicación del conocimiento	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
Mantener la calidad del trabajo producido	Media de las notas que cada estudiante individualmente recibió	Contribución individual en bytes
Capacidad crítica	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

Capacidad crítica. El indicador considerado para la capacidad crítica 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 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.

5. RESULTADOS DE LA INVESTIGACIÓN

5.2.2.2 Conclusiones

AMW es una herramienta de evaluación asistida que proporciona un análisis cualitativo interesante, pues al análisis cuantitativo proporcionado por StatMediaWiki, se podría añadir una información cualitativa que impidiera el engaño de los estudiantes. Por ejemplo, la contribución de un estudiante que copiase y pegase una gran cantidad de texto en una página del wiki sería considerada a efectos cuantitativos como una contribución de peso. Sin embargo, desde un punto de vista cualitativo, podría detectarse que la contribución había sido plagiada de internet con el fin de engañar al sistema.

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 también aparecerían.

5.3 Lenguajes de dominio

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 para obtener indicadores de los VLEs, mientras que en el apartado 5.3.2 se presenta el segundo DSL, orientado a los mundos virtuales.

5.3.1 EvalCourse (EVC) y SASQL

El VLE 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 VLE. Esta interacción será la base para los indicadores del desempeño de competencias genéricas.

5.3.1.1 Descripción

Para diseñar evaluaciones a partir de los indicadores del VLE se crea *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 [12]. De esta forma, los docentes pueden fácilmente obtener indicadores almacenados de la actividad en el VLE 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 indicadores solicitados. EvalCourse se comunica con el VLE para extraer la información del registro de actividad. EvalCourse¹ está basado en el IDE de la plataforma Eclipse, fue implementado utilizando Xtext [40] 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 VLE.

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 (*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 VLE. La conexión a la base de datos del curso se declara en la entidad *connection*.

¹<https://www.assembla.com/spaces/evalcourse>

5. RESULTADOS DE LA INVESTIGACIÓN

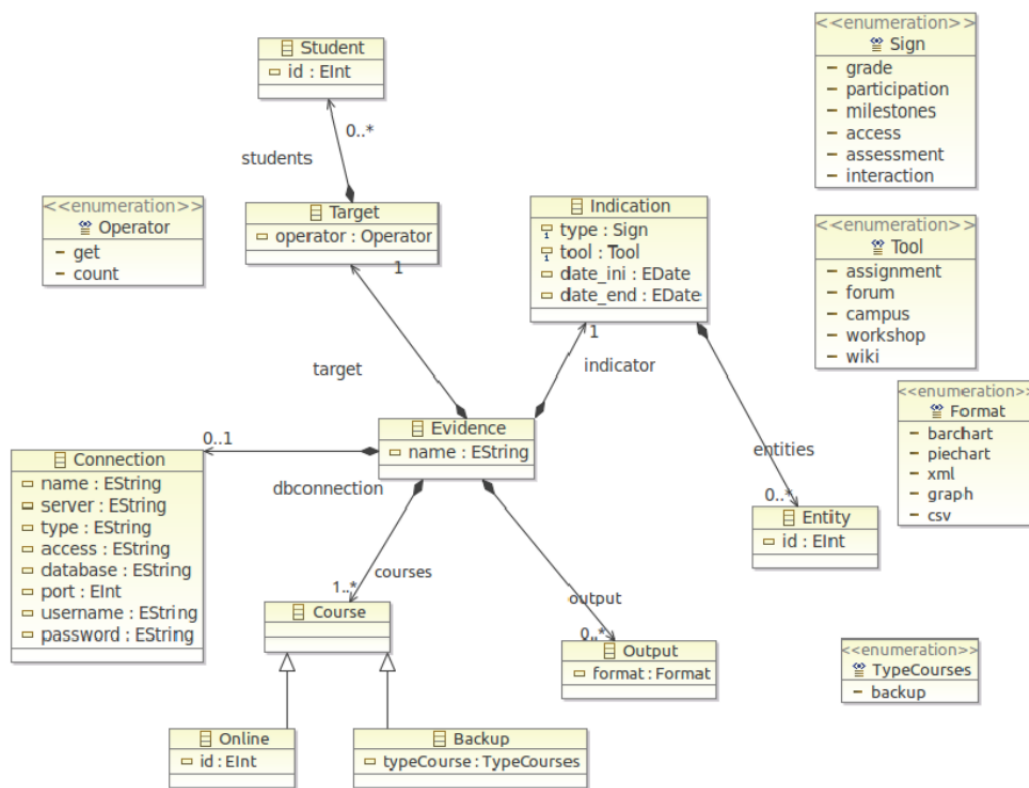


Figura 5.5: Metamodelo de SASQL

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   in assignment | forum | campus | workshop
6                   | wiki [list of ids]
7   between YYYY-MM-DD and YYYY-MM-DD
8   from course id | backup.
```

El objetivo es que EvalCourse pueda ser utilizado para cualquier VLE, pero esta primera versión ha sido implementada para funcionar en Moodle¹. Moodle es un sistema de gestión de aprendizaje de código abierto con más de 64.000 sitios registrados².

5.3.1.2 Ejemplo

Para evaluar las competencias genéricas el docente deberá definir los indicadores que serán extraídos de la actividad de cada estudiante en el VLE. Ilustraremos el

¹<https://moodle.org/>

²<https://moodle.net/stats>

5. RESULTADOS DE LA INVESTIGACIÓN

método a partir de un ejemplo de uso de EvalCourse y su ejecución con los foros del VLE. Los foros suelen ser una de las herramientas incorporadas por los VLEs 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 [46].

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 VLE, 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 cuantas 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. 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 periodo de tiempo particular 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*. 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.

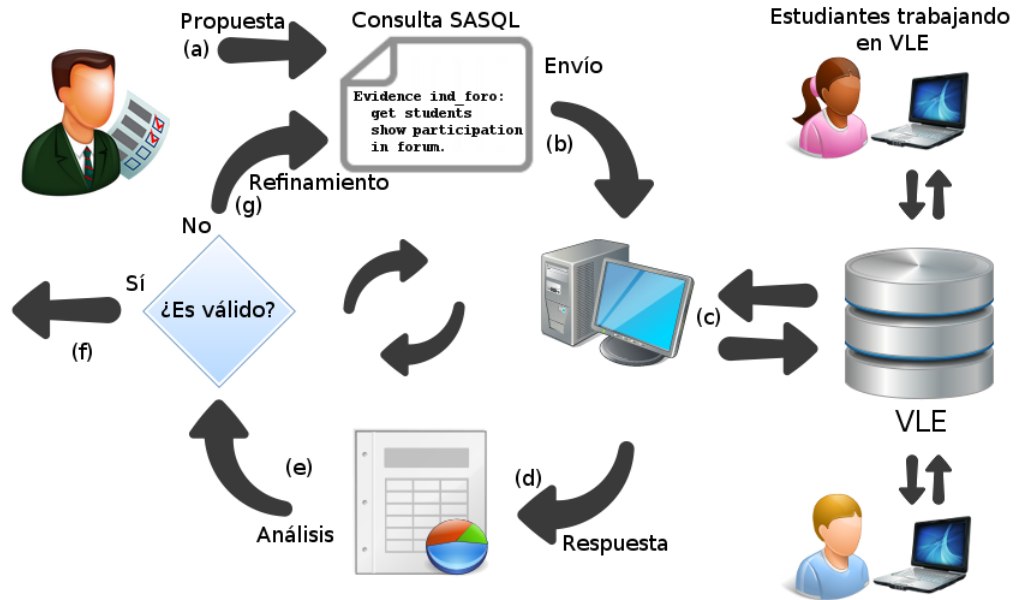


Figura 5.6: Ciclo de contraste de hipótesis utilizando EvalCourse

```

_____ Consulta SASQL participaciones foro _____
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

Tabla 5.2: 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

5. RESULTADOS DE LA INVESTIGACIÓN

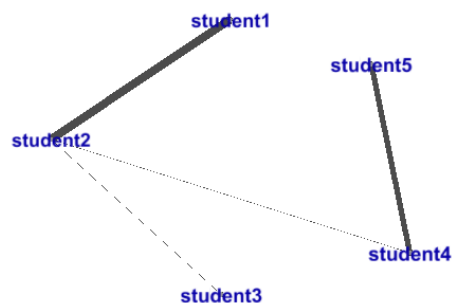


Figura 5.7: Interacción en el foro en un periodo de tiempo.

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 habilidades interpersonales si ha escrito al menos dos mensajes en el foro y ha interactuado con más de un compañero*. Para ello escribe la consulta SASQL interacciones foro.

```
_____ 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 [9]. Pero 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

Para la evaluación de esta competencia genérica se propone utilizar la participación en el foro. Al crear grupos de trabajo en el VLE 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

Para evaluar el liderazgo 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 estudiante

5. RESULTADOS DE LA INVESTIGACIÓN

que iniciara dos o más debates tendría una evaluación positiva en la competencia de liderazgo. 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.

```
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 esta competencia genérica. 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.
```

Pensamiento crítico

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

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 del pensamiento crítico, 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 nos permita utilizar este indicador utilizaríamos 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 estudiantes. El docente utilizaría la diferencia entre estos valores para valorar la competencia genérica.

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

Planificación y gestión del tiempo

Las tareas programadas en el VLE 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 esta se ha entregado a 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 se refiere. Pero con respecto a la planificación y gestión del tiempo de 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 utilizaríamos 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

5. RESULTADOS DE LA INVESTIGACIÓN

información el docente podrá evaluar la planificación y gestión del tiempo de sus estudiantes.

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

5.3.2 EvalSim (EVS) 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 en el aprendizaje de los estudiantes [17]. Lo mismo se puede decir de los entornos de aprendizaje como los mundos virtuales (Second Life, OpenCobalt, etc.) [55]. Esto se debe a que normalmente no se distribuyen bajo licencia libre y, por consiguiente, los docentes no pueden analizar las interacciones de los estudiantes o analizar su impacto en el aprendizaje y en los resultados de aprendizaje [29, 70].

Ante esta situación, en un trabajo previo desarrollamos nuestro propio mundo virtual, basado en OpenSim (software de código abierto) [18]. De esta forma se podría evaluar la competencia de la comunicación en una segunda lengua a partir de las interacciones llevadas a cabo por los estudiantes en el mundo virtual.

A continuación se explicará cómo a partir de la actividad generada por los estudiantes se obtendrán indicadores para el diseño de evaluaciones de competencias genéricas.

5.3.2.1 Descripción

Para este trabajo se creó VWQL (*Virtual Worl Query language*). VWQL es un DSL creado para obtener indicadores objetivos de OpenSim. EvalSim es el sistema que procesa las consultas de VWQL. Ha sido desarrollado bajo un enfoque MDE para modelar procesos para la obtención de los indicadores que se requieran. Fue también implementado utilizando Xtext [40] dentro del Eclipse Modeling Framework.

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 idioma, en cuyo caso únicamente muestra, de las palabras introducidas en el chat de texto, las palabras en dicho idioma.
- *sentences*: número de frases escritas en el chat de texto.
- *single*: número de frases compuestas por 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. RESULTADOS DE LA INVESTIGACIÓN

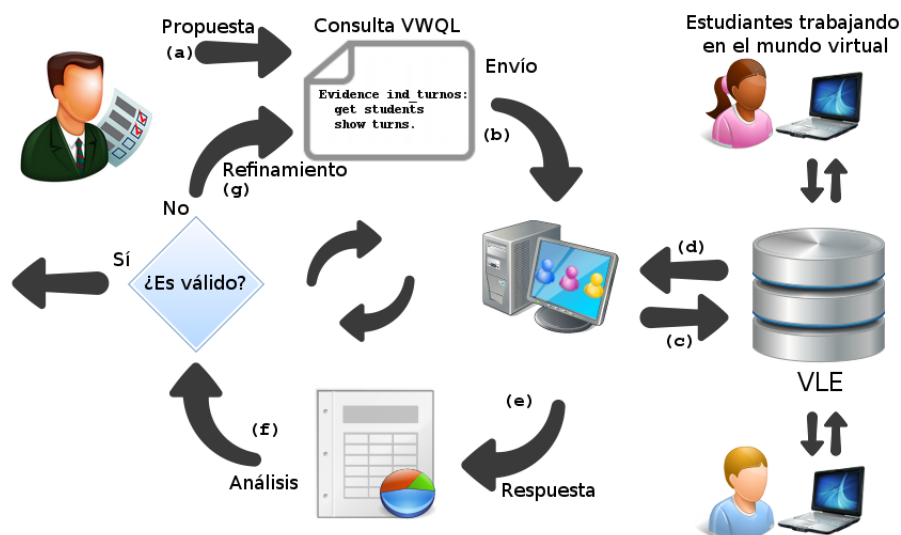


Figura 5.8: Ciclo de contraste de hipótesis con EvalSim

5.3.2.2 Ejemplo

En el juego los estudiantes interactúan entre ellos por medio de un chat. Las conversaciones quedan almacenadas en la base de datos de EvalSim. Utilizando el *ciclo de contraste de hipótesis* los docentes obtendrán indicadores del trabajo de sus estudiantes (figura 5.8).

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 idiomas. 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 en sus respuestas a fin de utilizar los recursos lingüísticos aprendidos en clase.

Tabla 5.3: 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 frases 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. 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 5.3. A la vista de los mismos, el docente considera que hay situaciones en las que una respuesta corta podría admitirse como válida, por lo que decide admitir como mucho 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 5.4: 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 5.4.

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 un segundo idioma*. No obstante, los indicadores podrían ser utilizados en la evaluación de otras competencias genéricas bajo el criterio del docente que diseña la evaluación.

Segundo idioma

- *Ritmo*: número de frases escritas por minuto. Según la experiencia desarrollada 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 tomarse como un indicador positivo o negativo dependiendo del 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.
- *Palabras solas*: número de frases 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 del segundo idioma.

Capacidad de aprender y mantenerse al día con el aprendizaje 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

Tabla 5.5: Resumen de publicaciones de los estudios de caso

Estudio de caso	Congreso	Revista
CS01	SPDECE 2012 ([11])	CHB 2014 ([75])
CS02	TEEM 2013 ([13])	IJEE 2015 ([9])
CS03	SIIE 2014 ([10])	IJEE 2015 ([9])
CS04	TEEM 2013 ([13])	IJEE 2015 ([9])
CS05	SIIE 2014 ([10])	IJEE 2015 ([9])
CS06	TEEM 2015 ([8])	JITR 2016 (In process)

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. A continuación se describen brevemente dichas experiencias, mientras que las publicaciones en las que se han presentado se resumen en la tabla 5.5 y se describen en el apartado 5.6.

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 público¹. La primera consistió en un proyecto de migración de la infraestructura 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 competencias del *trabajo en equipo*, *capacidad crítica*, la *comunicación*, la *aplicación del conocimiento*, la *capacidad para mantener la calidad del trabajo producido* y la *capacidad crítica de sus estudiantes*.

¹<http://wikis.uca.es/wikiASO>

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 las habilidades interpersonales y el liderazgo desempeñadas por los estudiantes en el foro de cada equipo.

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 del pensamiento crítico 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* 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

5. RESULTADOS DE LA INVESTIGACIÓN

herramienta EvalCourse para que pudiera obtener nuevo 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* en la Universidad de Cádiz durante el curso 2014/15. 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 *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 RR.PP.

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). La asignatura contaba con 5 estudiantes matriculados que 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 de la habilidad para comunicarse en un segundo idioma.

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 directrices marcadas por Oates [72] (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 de aprendizaje virtual
- *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 de aprendizaje virtual

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

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

5. RESULTADOS DE LA INVESTIGACIÓN

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 y viceversa. 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 Participantes

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.

5.5.2.1 Curso wiki innovación

Durante el curso 2014-15 se impartió a docentes de la Universidad de Cádiz un curso de innovación docente 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. El 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.

5.5.2.2 Taller Aulablog

Los días 6, 7 y 8 de julio del 2015 se celebró en Ubrique (Cádiz) el X encuentro Aulablog¹. 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.

5.5.2.3 Actuación avalada

Durante los cursos 2013-14 y 2014-15 dirigimos dos proyectos de tipo de actuación avalada dentro de los programas 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 RR.PP. Una vez finalizado cada curso se aplicó EvalCourse y se pusieron los indicadores obtenidos a disposición de los docentes para que lo utilizaran y propusiesen 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).

5.5.2.4 MediaWiki España

Se implicó a profesionales del ámbito de las TIC, y más en concreto de las wikis, pero no necesariamente ligados a la docencia. Mediante correo electrónico enviado a los miembros de MediaWiki España, se les explicó el método DBA, la herramienta EvalCourse y se les invita a completar el cuestionario.

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

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

5. RESULTADOS DE LA INVESTIGACIÓN

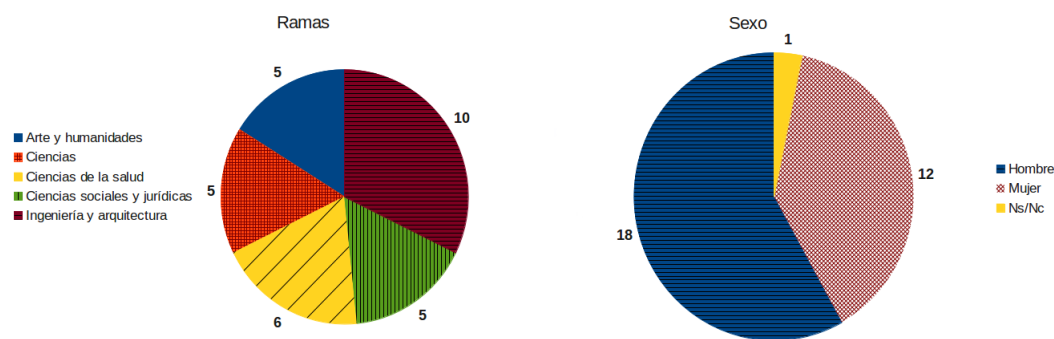


Figura 5.9: Distribución de los participantes por rama y sexo

5.5.2.5 Jornadas de Innovación Docente de la Universidad de Cádiz

En estas jornadas celebradas en el mes de marzo de 2016 se presentaron tanto el método DBA como el DSL que lo implementa. 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.

5.5.3.1 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 5.9. Además, en la tabla 5.6 se muestra un resumen de esta información.

5.5 Cuestionarios de evaluación

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 5.6: Resumen de distribución de participantes por rama y sexo

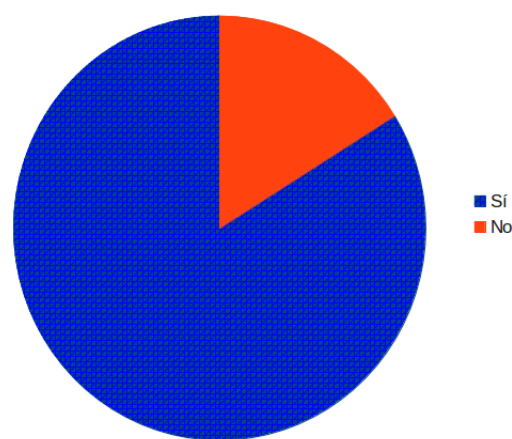


Figura 5.10: Respuestas de los encuestados a la pregunta sobre el método DBA

5.5.3.2 Evaluación del método DBA

Para evaluar el método se preguntó a los docentes si consideraban el método DBA adecuado para obtener automáticamente indicadores de entornos de aprendizaje virtual. A esta pregunta fueron 26 de los 31 docentes encuestados (84%) los que consideraron el método DBA adecuado (figura 5.10).

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 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 exacto de Fisher, no pudiéndose rechazar la hipótesis con un p-valor de 0,6142, muy superior

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 5.7: Consideración de idoneidad del método DBA por ramas

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 exacto 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 5.7.

5.5.3.2.1 Características del método DBA

Para continuar con la evaluación del método, se preguntó a los encuestados sobre su consideración acerca de en qué medida 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 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 5.8 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

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%	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 5.8: Evaluación de las características del método DBA

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.

5.5.3.3 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 de aprendizaje virtual. A esta pregunta también 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 do-

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 5.9: Consideración de idoneidad del DSL por ramas

cente 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 exacto 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 exacto 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 5.9.

5.5.3.3.1 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, 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 5.10 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 vo-

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%	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 5.10: Evaluación de las características del DSL

tado 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%).

5.5.3.4 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. Pero también se quiso determinar si existía relación entre la respuesta de cada do-

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 5.11: Consideración de la consulta como entendible o no por ramas

cente 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 exacto 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 exacto 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 5.11.

5.5.3.5 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 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 exacto

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 5.12: Consideración de los resultados como entendibles o no por ramas

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 exacto 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 5.12.

5.5.4 Resultados del cuestionario de indicadores

El cuestionario (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.

En total se consiguieron 51 cuestionarios completados, cifra que consideramos 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 5.11.

5. RESULTADOS DE LA INVESTIGACIÓN

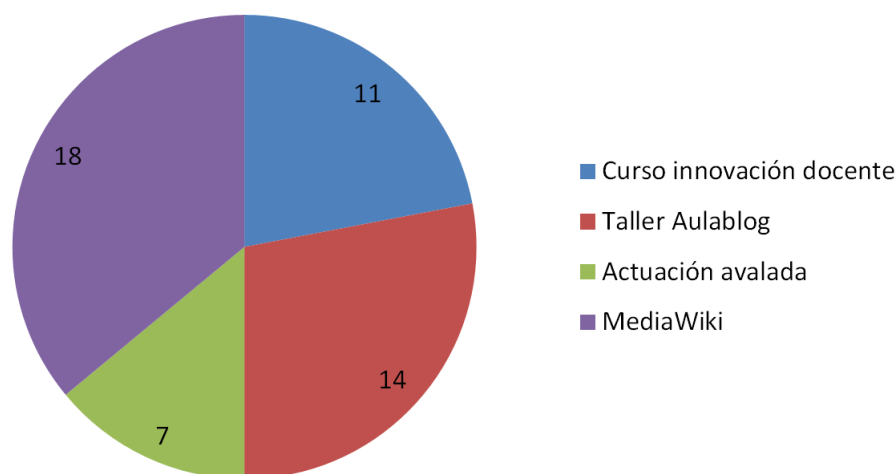


Figura 5.11: Distribución de los participantes en el cuestionario

5.5.4.1 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 5.13 se muestra la distribución de los participantes en base a estos atributos.

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 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

5.5 Cuestionarios de evaluación

	POBLACIÓN	Curso in- novación docente	Taller Aulablog	Actuación avalada	Media Wiki	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 5.13: Perfil de los participantes que participaron en el cuestionario

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

siguiente apartado.

5.5.4.2 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 del *trabajo en equipo*, la *planificación y gestión del tiempo* y el *liderazgo*. El resumen de las respuestas dadas puede verse en la tabla 5.14.

Vemos que la tendencia de las respuestas de los participantes es diferente para cada competencia. Mientras que para evaluar el *trabajo en equipo* podríamos decir que hay un empate entre el ‘sí’ y el ‘no’ (26 y 25 respectivamente), en la competencia de la *planificación y gestión del tiempo* gana ampliamente el ‘sí’ (40 a 11) y en la competencia del *liderazgo* gana, aunque no tan ampliamente, el no (17 a 34). Como puede verse las cifras son diferentes para cada competencia, lo que viene a

5. RESULTADOS DE LA INVESTIGACIÓN

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. Para abordar ahora la independencia de las respuestas de los participantes en base a su perfil y sus conocimientos técnicos vamos a utilizar la prueba χ^2 (*chi cuadrado*) de Pearson.

5.5.4.2.1 Independencia de las respuestas

Para probar la independencia de las variables que intervienen en el cuestionario se utilizará la prueba χ^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 competencia de **trabajo en equipo**
2. 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**
3. Los **conocimientos de programación** son independientes de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **liderazgo**
4. 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**
5. 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**

6. El **perfil** del individuo es independiente de que el individuo considere que les son válidos los indicadores extraídos para medir la competencia de **liderazgo**

Se construyeron las tablas de contingencia, se calcularon las frecuencias esperadas y se obtuvieron los diferentes valores de χ^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 estudio completo puede verse en la sección A.2 del apéndice A.

5.5.4.2.2 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.2.4.3. A continuación se resumen las justificaciones para cada competencia.

Trabajo en equipo

Los participantes que consideran que sí se puede evaluar el trabajo en equipo 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 que demandan que para medir un trabajo en equipo 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

En esta competencia el indicador que más defienden los participantes para evaluar el trabajo en equipo 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.

5. RESULTADOS DE LA INVESTIGACIÓN

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

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 qué 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.

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 el liderazgo. Se propone como alternativa opuesta que un indicador de liderazgo 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 los proyectos de actuación avalada 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 5.15 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, seguida de la planificación y gestión del tiempo. En la sección B.2 del apéndice B pueden verse al

5.5 Cuestionarios de evaluación

Tabla 5.15: Cuadro de indicadores considerados para cada competencia

Competencia	Accesos	Foros	Actividades	Wiki	Talleres
Trabajo en equipo	1	2	2	4	0
Planificación y gestión del tiempo	5	2	6	5	0
Razonamiento crítico y autocrítico	0	4	2	1	0
Capacidad para identificar, plantear y resolver problemas	0	5	4	2	1
Habilidades de interacción interpersonal	2	6	0	3	1
Trabajo autónomo	6	3	5	7	1

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 entornos de aprendizaje virtual y consideran el DSL útil para diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de dichos entornos.

5. RESULTADOS DE LA INVESTIGACIÓN

A las encuestas se les pasó el test exacto 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 estadístico 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 entendibles 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 exacto 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 tan amplia como en los casos anteriores. Sobre todo es llamativo lo que ocurre con los resultados, donde todos los docentes del area 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 de aprendizaje virtual
- *H2*: El DSL permite a los docentes diseñar y contrastar estrategias de evaluación a partir de los registros de actividad de los entornos de aprendizaje virtual

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 satisfacen las características de la *adaptabilidad*, *sistematicidad* y la *flexibilidad*. Las características de la *objetividad* y la *fiabilidad* también tienen un mayor

porcentaje de votos a favor que en contra, aunque con menor porcentaje (en torno al 45%).

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. En los ejemplos mostrados se proponen ejemplos de indicadores, y según el criterio de cada participante en el cuestionario, unos son más válidos que otros. Para la competencia del *trabajo en equipo* prácticamente hay un empate entre el ‘sí’ (26) y el ‘no’ (25). Mientras que en la *planificación y gestión del tiempo* domina el ‘sí’ (40 a 11), en la competencia del *liderazgo* 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, partiendo de los diferentes perfiles de los participantes en la encuesta se demuestra que sus respuestas son independientes tanto de su perfil docente como de su perfil más o menos técnico. 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 estos los únicos que acepten los diseños, o que esté más cercano a un docente de perfil universitario. El establecimiento de una serie de hipótesis nulas y la prueba χ^2 de Pearson así lo han demostrado.

Hay que comentar que el hecho de que la hipótesis nula número 3 fuese más difícil de demostrar tiene una explicación. La mayoría de los usuarios que dijo ‘sí’ a la posibilidad de evaluar la competencia de liderazgo fueron los participantes

5. RESULTADOS DE LA INVESTIGACIÓN

en el curso de innovación 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 cuestionario complementario utilizado en la actuación avalada 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.

5.6 Contribuciones

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

Revistas

- Artículo en la revista *Computers in Human Behavior* (2014), titulado “Scalability of assessments of wiki-based learning experiences in higher education” [75].
- Artículo en la revista *International Journal of Engineering Education* (2015), titulado “A Domain Specific Language for Online Learning Competence Assessments” [9].

Congresos

- Artículo presentado en el congreso IX Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos (SPDECE 2012), titulado “Qualitative assessment of wiki-based learning processes” [11].
- Artículo presentado en el I International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM 2013), titulado “A generative computer language to customize online learning assessments” [13].
- Artículo presentado en el International Symposium on Computers in Education (SIIE 2014), titulado “Domain-driven competence assessment in virtual learning environments. Application to planning and time management skills” [10].
- Artículo presentado en el III International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM 2015), titulado “A Domain Specific Language to retrieve objective indicators for foreign language learning in virtual worlds” [8].
- Artículo presentado en el III Congreso Internacional Sobre Aprendizaje, Innovación y Competitividad (CINAIC 2015), titulado “Evaluación del trabajo individual y grupal en un wiki” [85].

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

Winston Churchill

CAPÍTULO

6

Conclusiones y trabajo futuro

En este capítulo se presentan las conclusiones de este trabajo de investigación. Además se comentan los trabajos futuros que se pueden realizar para continuar esta línea de investigación.

6.1 Conclusiones

En esta tesis doctoral se ha presentado el método DBA y se ha propuesto su utilización para la evaluación de competencias genéricas. El método DBA tiene su origen en la metodología DBR y se basa en el diseño de evaluaciones a partir de indicadores de la actividad de los estudiantes generada en los entornos virtuales de apoyo a la docencia. El diseño de evaluaciones se realiza dentro del ciclo de contraste de hipótesis. De esta forma, un docente que quiera realizar una evaluación formula un diseño (hipótesis de evaluación) a partir de los indicadores del entorno virtual, el sistema devuelve los resultados de aplicar esta fórmula y el docente evalúa su validez para la evaluación de la competencia. Mientras no le sean válidos, el docente seguirá rediseñando la fórmula de evaluación hasta alcanzar un diseño que le satisfaga o hasta que tenga que descartar los resultados.

6. CONCLUSIONES Y TRABAJO FUTURO

Para poner en práctica el método DBA se han propuesto herramientas de tipo DSL. En primer lugar se desarrolló SASQL, un DSL para obtener indicadores de la actividad de los estudiantes del VLE y EvalCourse, el software que interpreta las consultas escritas en SASQL. En segundo lugar se desarrolló VWQL, DSL que en este caso obtiene indicadores de la actividad de los estudiantes de los mundos virtuales y EvalSim, el software que interpreta las consultas escritas en VWQL.

Se han desarrollado seis estudios de caso entre los cursos 2010-11 y 2014-15 en el que los docentes han utilizado los DSL que implementan el método DBA para la evaluación de competencias genéricas. Estas experiencias han servido para obtener retroalimentación de los docentes que han participado. Además, estos estudios de caso se han presentado en varios congresos y revistas, pasando por procesos de revisión ciega por pares.

De entre los participantes en estos estudios de caso y otros contextos educativos se ha llevado a cabo una encuesta para llevar a cabo la evaluación del método. En primer lugar, con esta encuesta se ha dado validez al método para ser utilizado en la evaluación de diferentes competencias genéricas. Los indicadores han sido aceptados como válidos para evaluar competencias genéricas por un porcentaje alto de docentes para cada competencia genérica. Además, aquellos docentes que no consideraron la evaluación de alguna competencia lo hicieron porque consideraban que el indicador debería ser completado con más información. Asimismo, mediante la formulación de diferentes hipótesis nulas se ha demostrado que las respuestas de los participantes en la encuesta son independientes tanto de los conocimientos de programación (nulos o básicos frente a medios o avanzados) como de la relación de los participantes con la docencia (profesorado universitario, profesorado no universitario o no profesorado relacionado con las wikis).

Por tanto, de los resultados de la evaluación realizada, se considera que la metodología DBA parece ser una propuesta eficaz para abordar la evaluación de competencias genéricas, ya que aporta indicadores objetivos del trabajo de los estudiantes en las herramientas virtuales, indicadores que han sido aceptados por los docentes para la evaluación de competencias genéricas y cuya obtención mediante las herramientas proporcionadas no les requiere conocimientos informáticos avanzados.

6.2 Trabajo futuro

Durante la realización de este trabajo se han superado los retos que se plantearon inicialmente. Sin embargo, durante la realización de la misma, 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 serían:

- Añadir indicadores basados en la aplicación de técnicas del análisis de las redes sociales (SNA) 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 6.1. En dicho grafo, el nodo del medio es el profesor. Puede verse como al principio el profesor 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 arista que une normalmente dos nodos periféricos representa los estudiantes que colaboraron juntos en una página. El tamaño de los nodos representa la cantidad de trabajo realizado por cada estudiante.
- Integrar los DSL con los propios entornos virtuales de aprendizaje. En este momento las herramientas virtuales (VLE, wiki o mundo virtual) funcionan por su lado. Los DSL se conectan a ellos de algún modo y se obtienen los indicadores. Sería interesante poder integrar el DSL en dichas herramientas dentro del perfil de supervisor que se otorga a los docentes en muchas de estas herramientas virtuales.
- Establecer una biblioteca que relacione actividades a realizar por los estudiantes en los entornos de aprendizaje, indicadores que se generan y competencias genéricas que se evaluarían con dichos indicadores.

6. CONCLUSIONES Y TRABAJO FUTURO



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

APÉNDICE



Cuestionario

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

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 A.1: ¿Sería capaz de valorar si los tres estudiantes han participado activamente en el wiki?

- Sí
- No

Evidence Participacion :
get students
show participation
in wiki .

Consulta A.1: Consulta participación de los estudiantes en el wiki

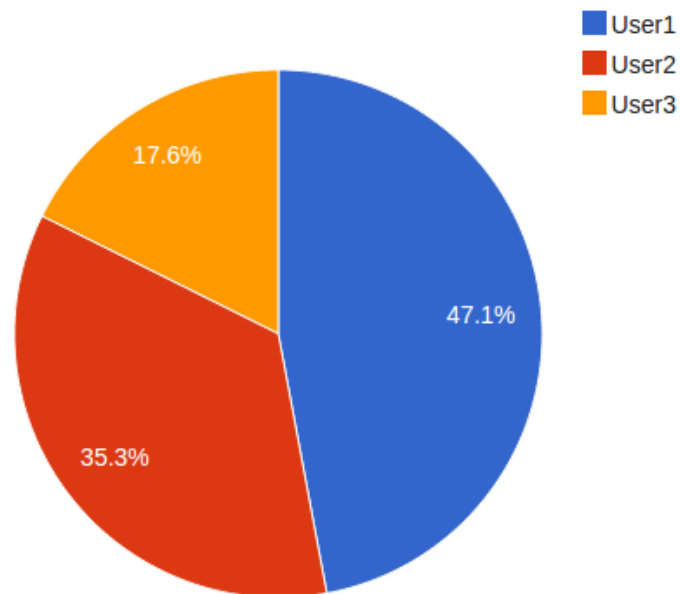


Figura A.1: Participación de los estudiantes en la página del wiki

A. CUESTIONARIO

C.2 Consulta de indicadores 2

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

- Sí
- No

Evidence Contribuciones :
get students
show interaction
in wiki .

Consulta A.2: Consulta contribuciones de los estudiantes en el wiki

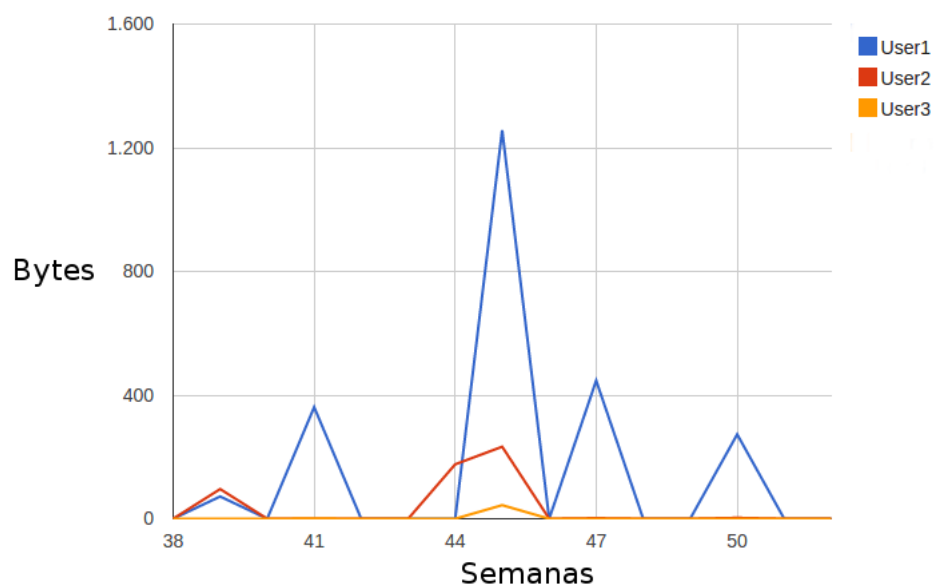


Figura A.2: Contribuciones de los estudiantes a la página del wiki

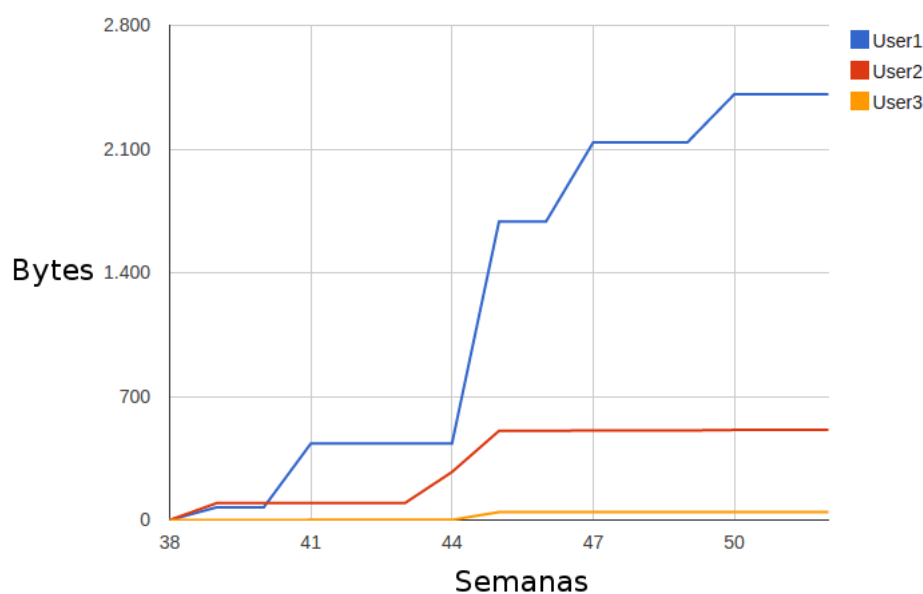


Figura A.3: Evolución del contenido de una página del wiki

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

- 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 Poblaciones

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.

A.2.1.1 Curso innovación docente

Profesorado de la Universidad de Cádiz que participó en el curso de innovación 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: 11

A.2.1.2 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

- Número de participantes: 14

A.2.1.3 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 sus cursos virtuales que favoreciesen la actividad de los estudiantes en el curso para aplicar después EvalCourse. Se hizo especial hincapié en la inclusión de wikis en sus cursos virtuales.
- Número de participantes: 7

A.2.1.4 MediaWiki

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

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

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

A.2.2 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

A.2 Resultados

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

	POBLACIÓN	Curso in- novación docente	Taller Aulablog	Actuación avalada	Media Wiki	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
Profesorado	Universitario	11	2	7	1	21 (41%)
	No universitario	0	12	0	0	12 (23%)
	No profesorado	0	0	0	18	18 (35%)

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

(35%), pero que tiene un profundo conocimiento en el uso de wikis, ya que son componente de MediaWiki 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.2.3 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 par-

A. 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

ticipación 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.2.4 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.2.4.1 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

A.2 Resultados

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

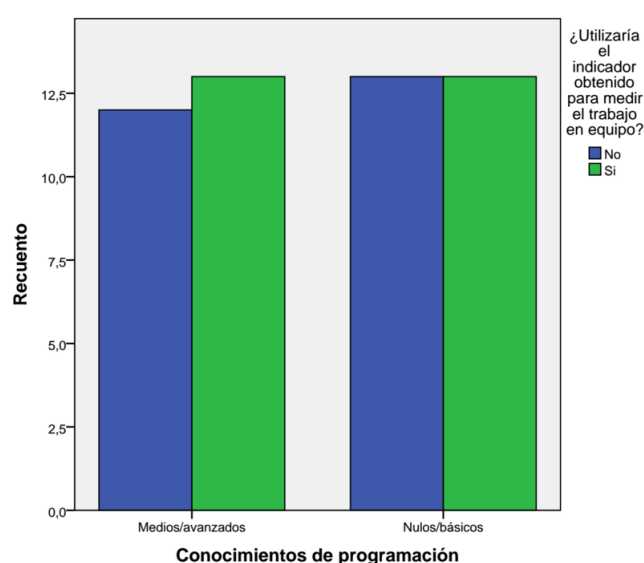


Figura A.4: Utilizaría el indicador para evaluar el trabajo en equipo

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.

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:

A. CUESTIONARIO

*H₀: 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 (χ_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.

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

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:

*H₀: 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 (χ_0^2):

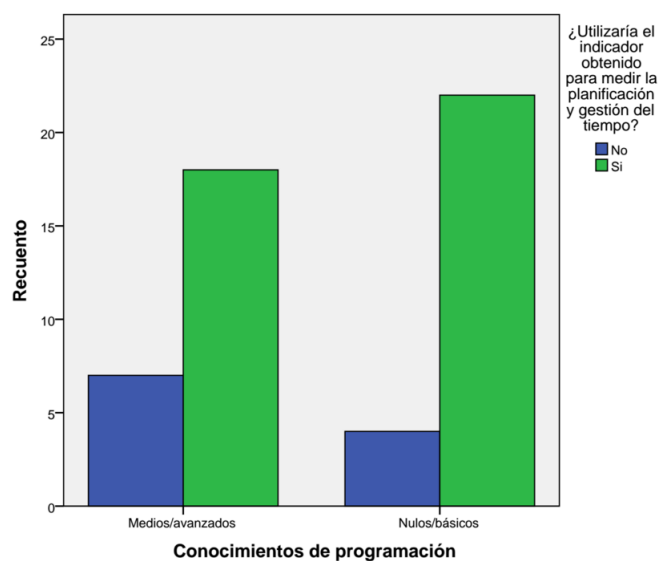


Figura A.5: Utilizaría el indicador para evaluar la planificación y gestión de tiempo

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

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.

A. CUESTIONARIO

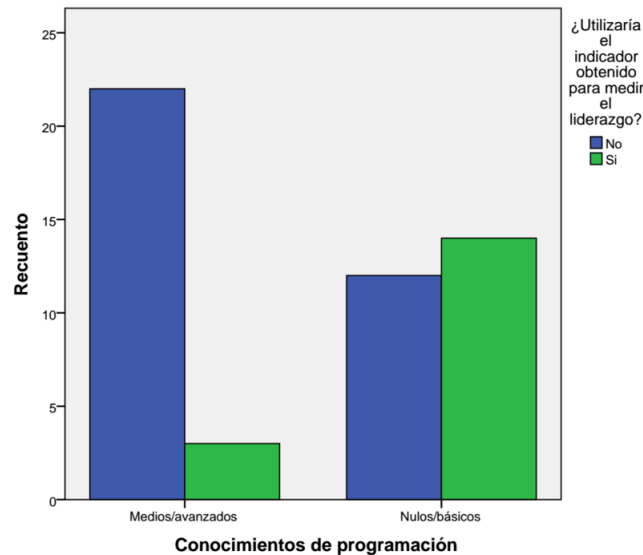


Figura A.6: Utilizaría el indicador para evaluar el liderazgo

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 (χ_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.

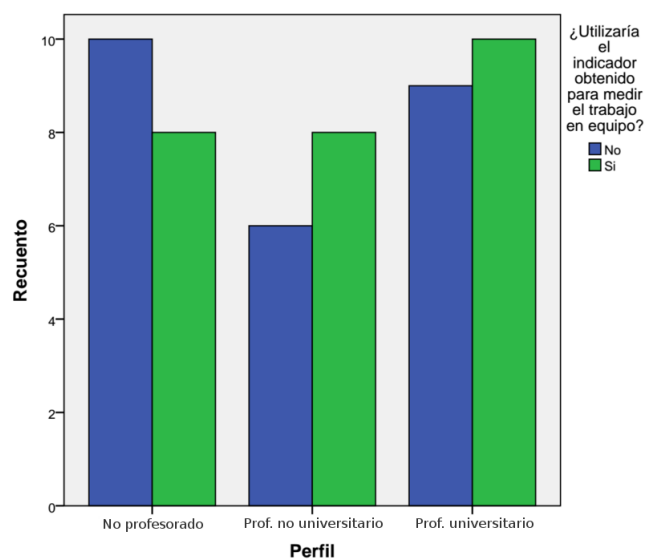


Figura A.7: Utilizaría el indicador para evaluar el trabajo en equipo

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

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.2.4.2 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.

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:

A. CUESTIONARIO

*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 (χ_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.

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

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 (χ_0^2):

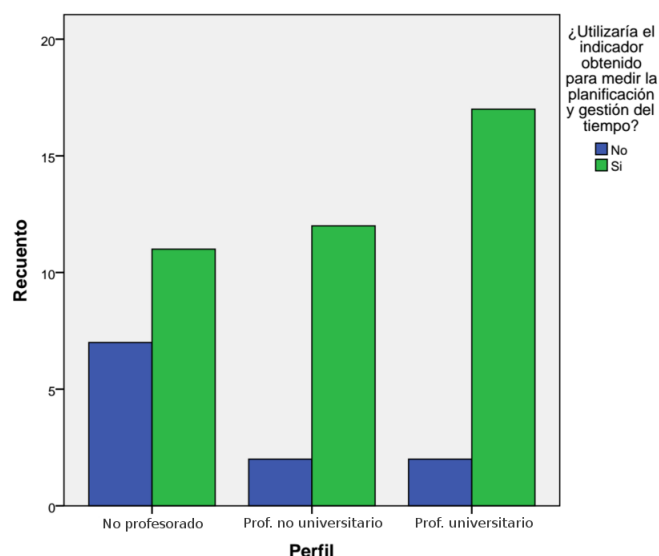


Figura A.8: Utilizaría el indicador para evaluar la planificación y gestión de tiempo

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

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:

A. CUESTIONARIO

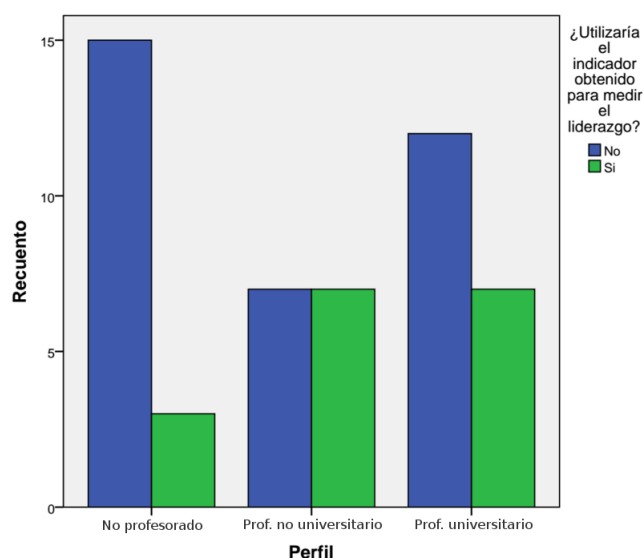


Figura A.9: Utilizaría el indicador para evaluar el liderazgo

*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 **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 (χ_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.2.4.3 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
- 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

A. CUESTIONARIO

- 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
- 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 interaccionado 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

A. CUESTIONARIO

- 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
- 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

A. CUESTIONARIO

- 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
- 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.

A. CUESTIONARIO

- 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
- 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

A. CUESTIONARIO

- 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
- Incidiendo 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 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 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 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.

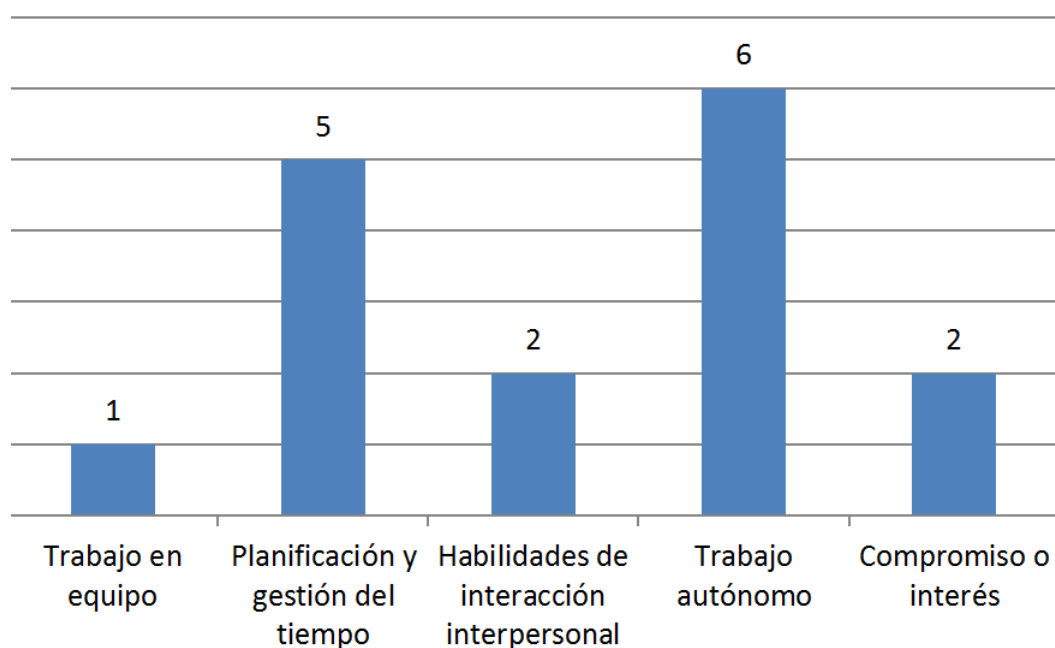


Figura B.1: Docentes que considerarían los indicadores de accesos para evaluar cada competencia genérica

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 el grado de Publicidad y Relaciones Públicas, un grado de la Facultad de Ciencias Sociales y de la Comunicación 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 este indicador, conociendo el número de días que se imparte la

B. CUESTIONARIO COMPLEMENTARIO ACTUACIÓN AVALADA

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...

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.

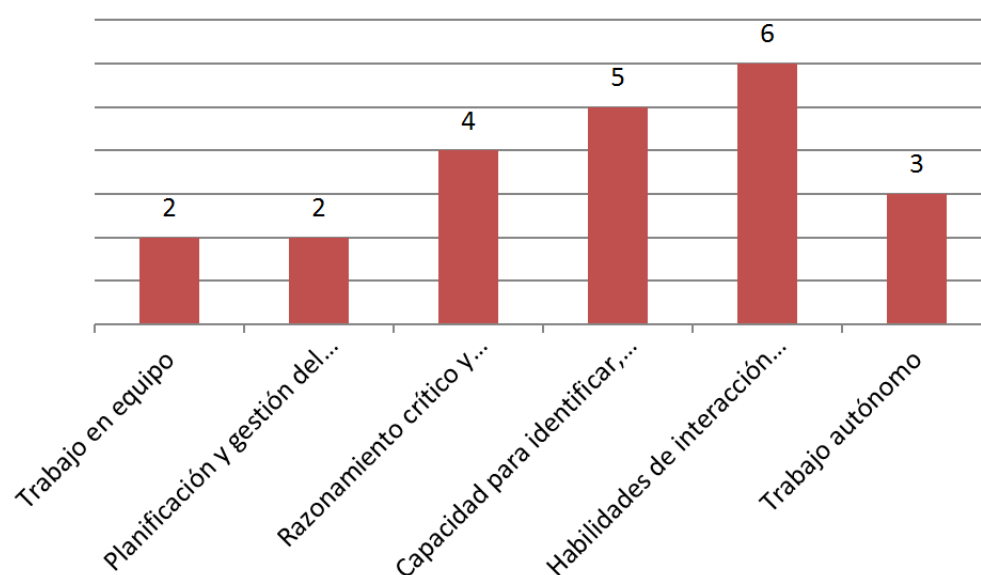


Figura B.2: Docentes que considerarían los indicadores de los foros para evaluar cada competencia genérica

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 (abcisas). 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 es

B. CUESTIONARIO COMPLEMENTARIO ACTUACIÓN AVALADA

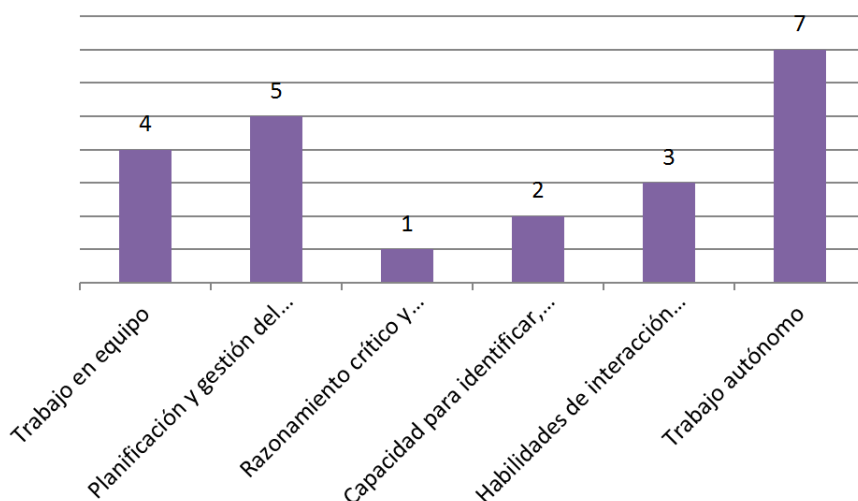


Figura B.3: Docentes que considerarían los indicadores de los wikis para evaluar cada competencia genérica

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.

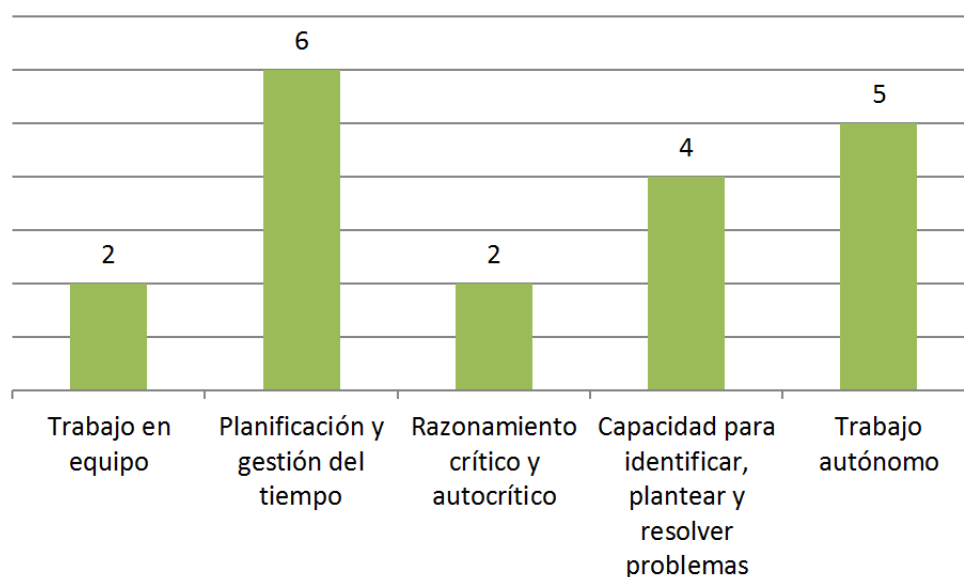


Figura B.4: Docentes que considerarían los indicadores de las actividades para evaluar cada competencia genérica

- 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 su caso lo 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 entornos de aprendizaje virtual?

- 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 de aprendizaje virtual?

- 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:

```

1 Evidence interaccion_foro:
2     get students
3     show interaction
4     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:

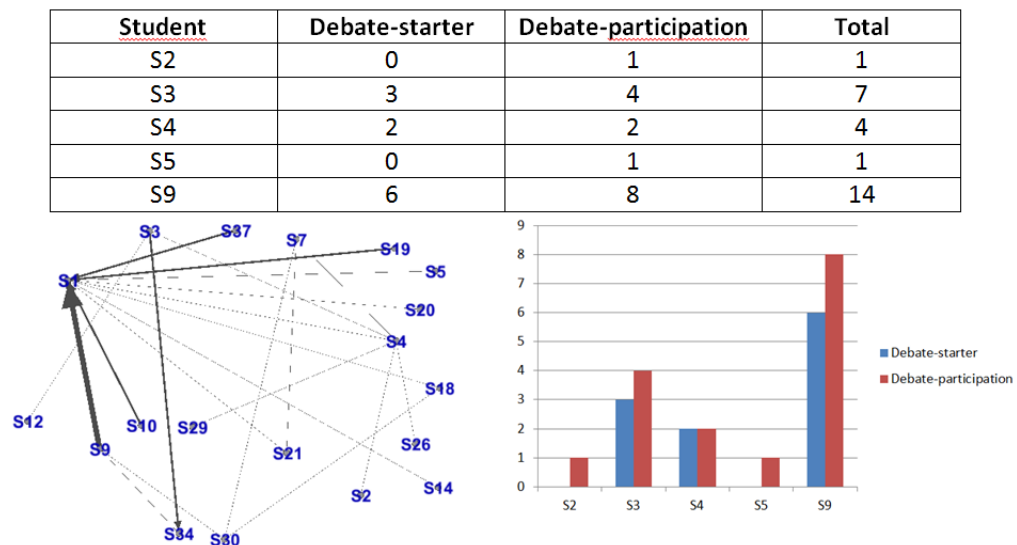


Figura C.1: Resultados consulta

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 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 exacto 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 exacto 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.

C.2.2.1 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.

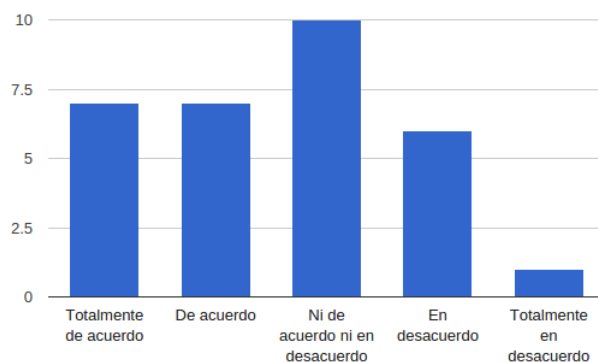


Figura C.2: Resumen de respuestas para la característica de la objetividad

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.

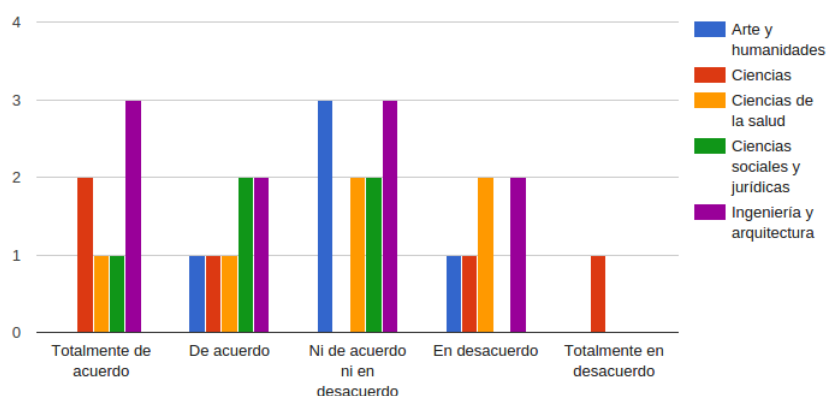


Figura C.3: Resumen de respuestas para la característica de la objetividad por rama

Adaptabilidad

El resumen de respuestas para la característica de la adaptabilidad puede verse en las figuras C.4 y C.5.

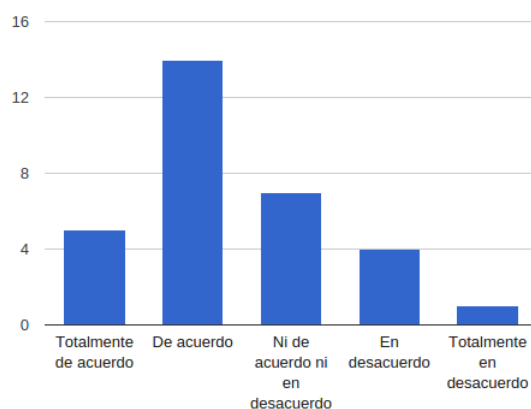


Figura C.4: Resumen de respuestas para la característica de la adaptabilidad

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.

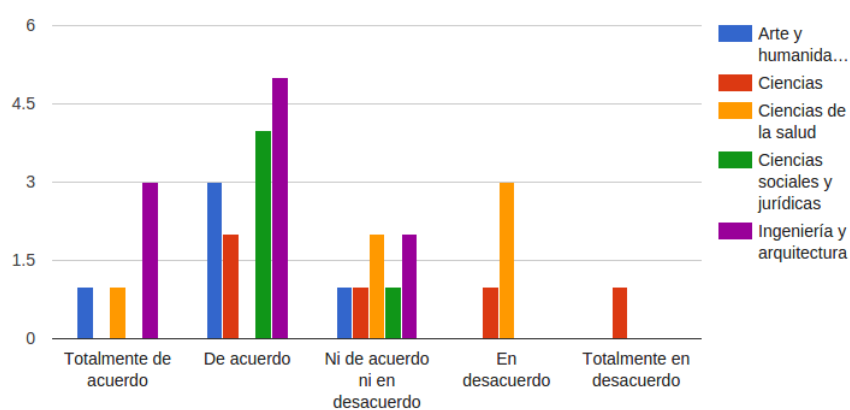


Figura C.5: Resumen de respuestas para la característica de la adaptabilidad por rama

C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

Sistematicidad

El resumen de respuestas para la característica de la sistematicidad puede verse en las figuras C.6 y C.7.

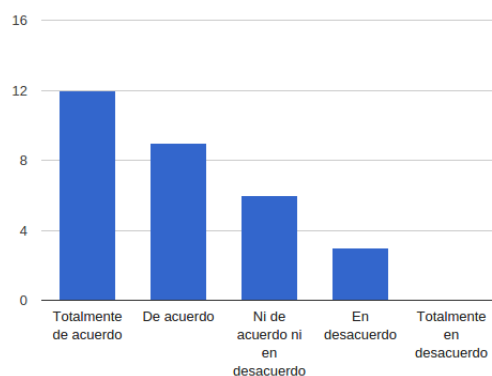


Figura C.6: Resumen de respuestas para la característica de la sistematicidad

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.

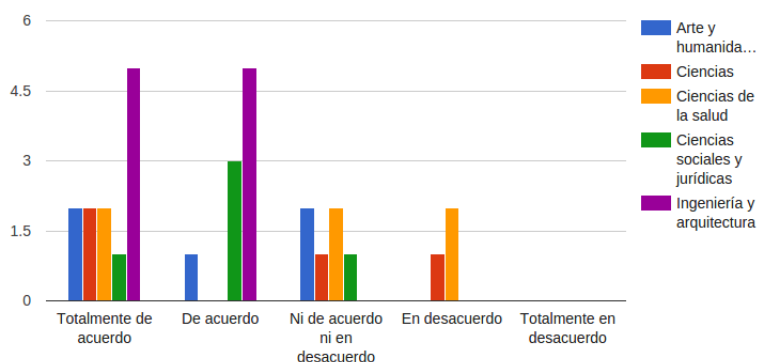


Figura C.7: Resumen de respuestas para la característica de la sistematicidad por rama

Flexibilidad

El resumen de respuestas para la característica de la flexibilidad puede verse en las figuras C.8 y C.9.

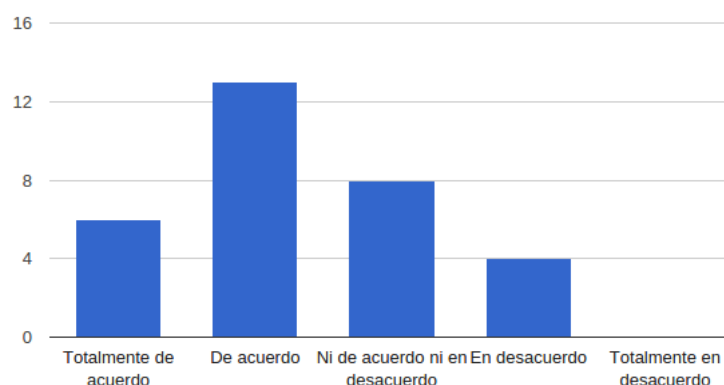


Figura C.8: Resumen de respuestas para la característica de la flexibilidad

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.

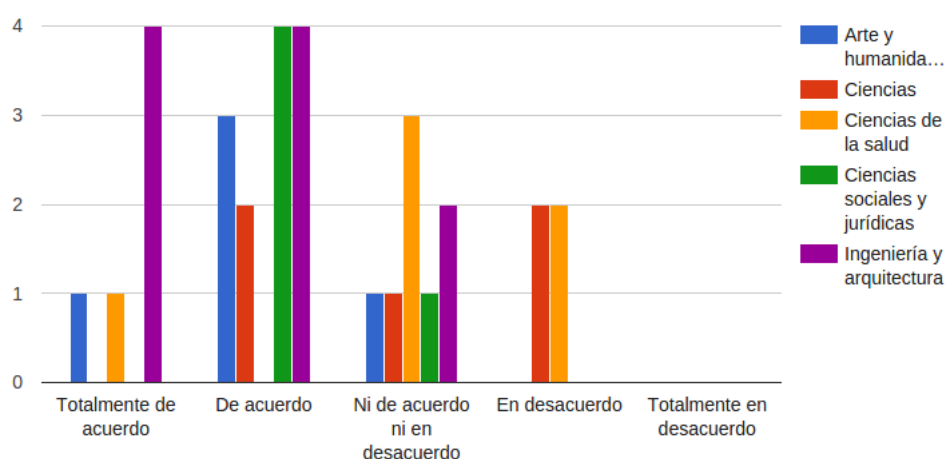


Figura C.9: Resumen de respuestas para la característica de la flexibilidad por rama

C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

Fiabilidad

El resumen de respuestas para la característica de la fiabilidad puede verse en las figuras C.10 y C.11.

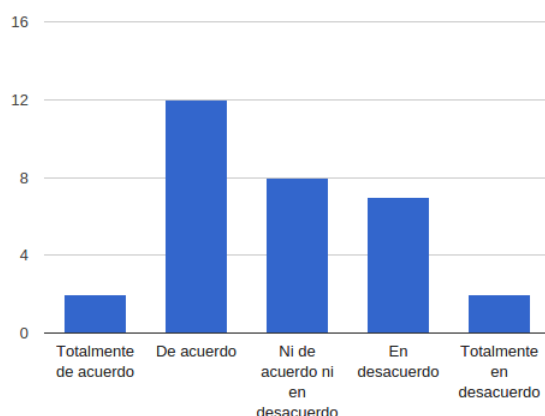


Figura C.10: Resumen de respuestas para la característica de la fiabilidad

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 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.

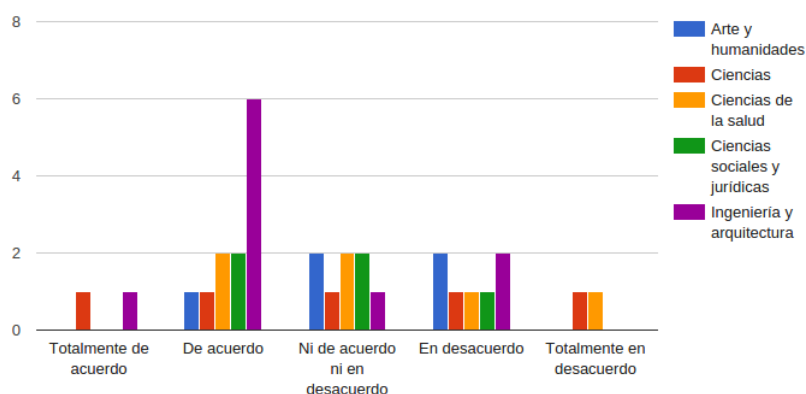


Figura C.11: Resumen de respuestas para la característica de la fiabilidad por rama

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_0 : 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 exacto 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 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 exacto 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.

C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

C.2.3.1 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.

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.

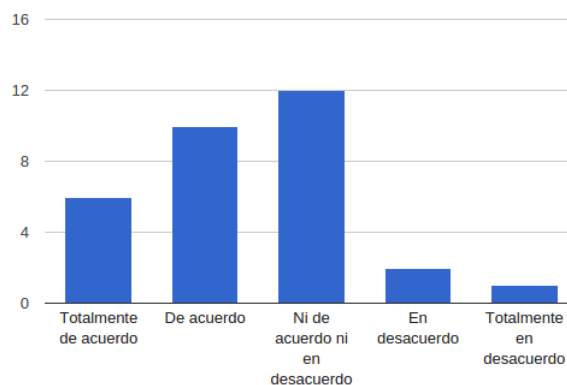


Figura C.12: Resumen de respuestas para la característica de la facilidad de aprendizaje

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.

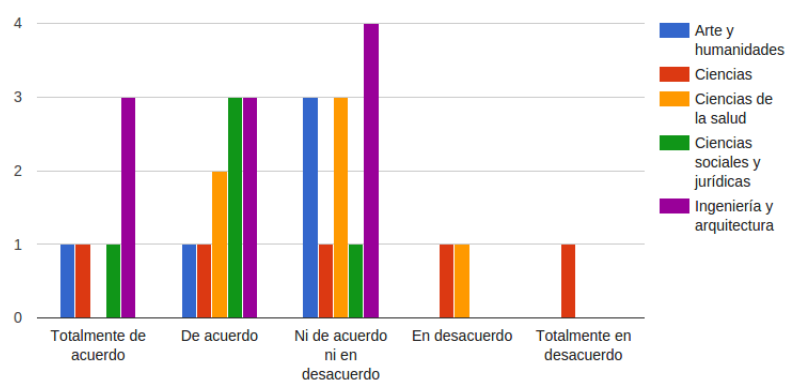


Figura C.13: Resumen de respuestas para la característica de la facilidad de aprendizaje por rama

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.

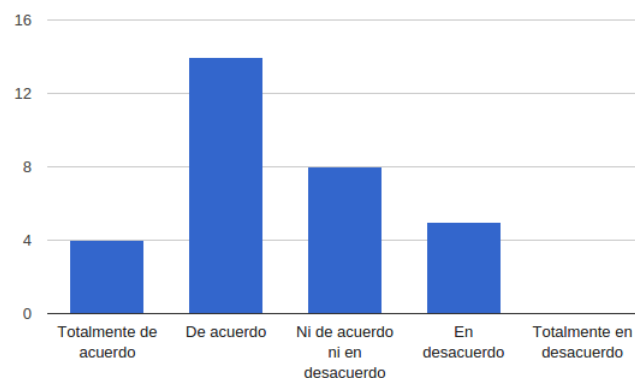


Figura C.14: Resumen de respuestas para la característica de la eficacia

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.

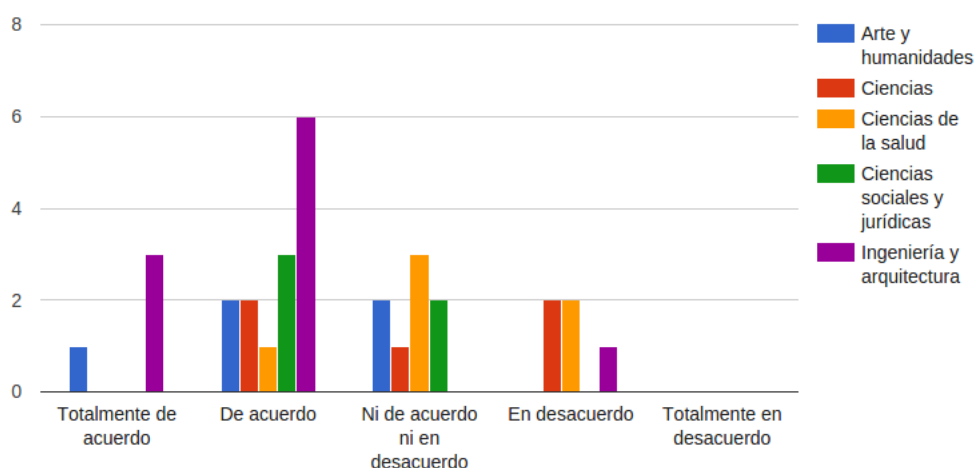


Figura C.15: Resumen de respuestas para la característica de la eficacia por rama

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.

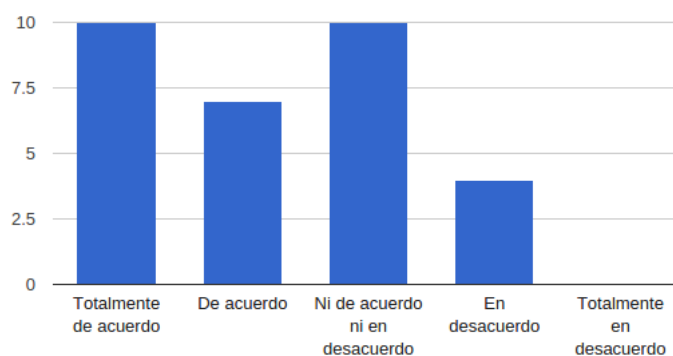


Figura C.16: Resumen de respuestas para la característica del ahorro del tiempo

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 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.

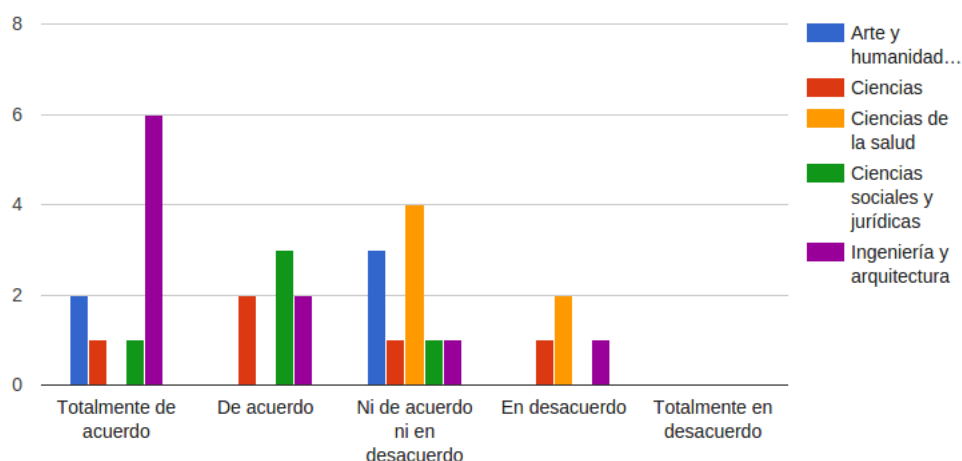


Figura C.17: Resumen de respuestas para la característica del ahorro del tiempo

C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

Escalabilidad

El resumen de respuestas para la característica de la escalabilidad del DSL puede verse en las figuras C.18 y C.19.

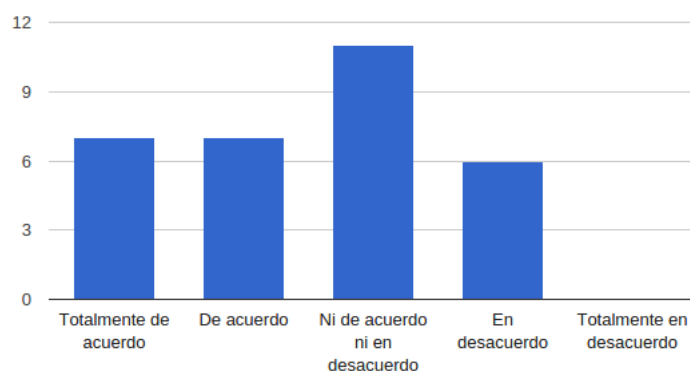


Figura C.18: Resumen de respuestas para la característica de la escalabilidad

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.

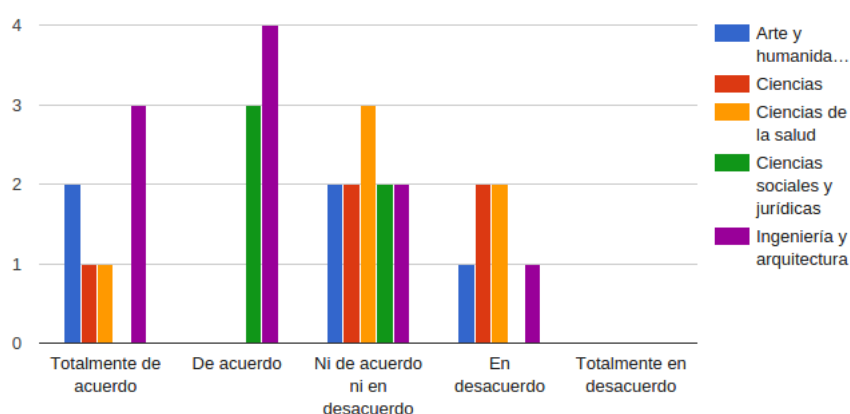


Figura C.19: Resumen de respuestas para la característica de la escalabilidad por rama

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.

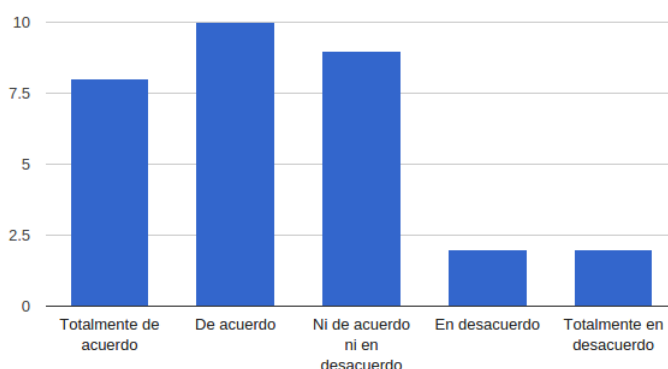


Figura C.20: Resumen de respuestas para la característica de la capacidad de reutilización

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.

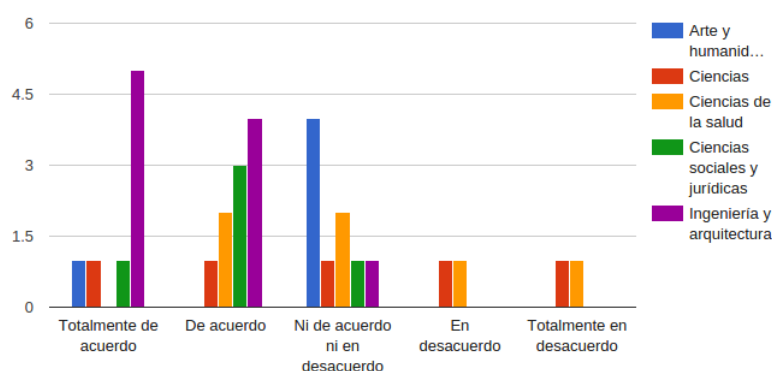


Figura C.21: Resumen de respuestas para la característica de la capacidad de reutilización por rama

C. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

Fiabilidad

El resumen de respuestas para la característica de la fiabilidad del DSL puede verse en las figuras C.22 y C.23.

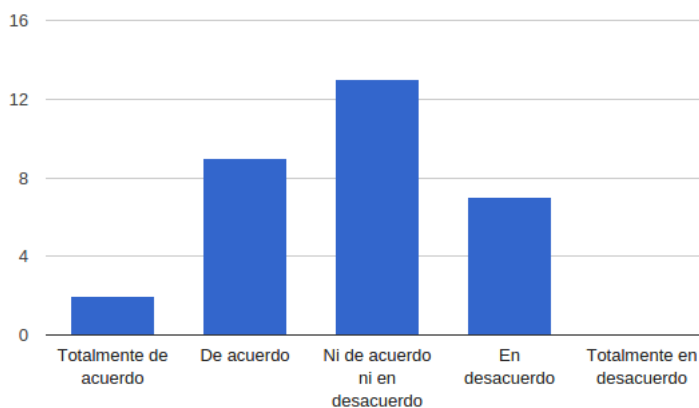


Figura C.22: Resumen de respuestas para la característica de la fiabilidad

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.

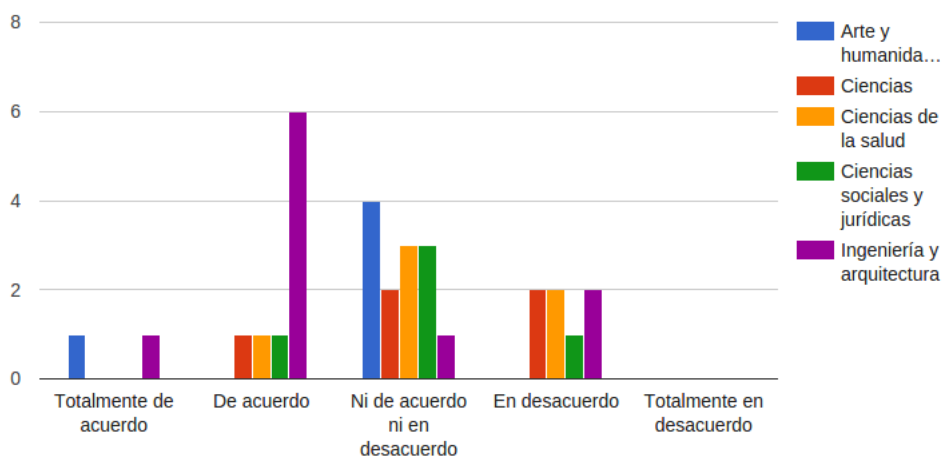


Figura C.23: Resumen de respuestas para la característica de la fiabilidad por rama

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

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 exacto 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 exacto 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. EVALUACIÓN DEL MÉTODO DE EVALUACIÓN (DBA) Y DEL LENGUAJE PARA DISEÑAR EVALUACIONES (SASQL)

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 exacto 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

Para contrastar la hipótesis se utilizó el test exacto 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.

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. 38, 40, 41, 49
- [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. 16
- [3] T. Anderson and J. Shattuck. Design-based research a decade of progress in education research? *Educational researcher*, 41(1):16–25, 2012. 7, 73, 74
- [4] 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. 44, 45, 46, 50, 64
- [5] 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>. Accessed: 2016-01-11. 2
- [6] 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. 38, 40, 41, 51

BIBLIOGRAFÍA

- [7] 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. 43, 49
- [8] 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. 110, 133
- [9] 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, 110, 132
- [10] 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. 110, 133
- [11] 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. 110, 133
- [12] 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. 95
- [13] A. Balderas, I. Ruiz-Rube, M. Palomo-Duarte, and J. M. Dodero. A generative computer language to customize online learning assessments. In *Pro-*

- ceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, pages 141–147. ACM, 2013. 110, 133
- [14] M. Bedek, S. A. Petersen, and T. Heikura. From behavioral indicators to contextualized competence assessment. In *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, pages 277–281. IEEE, 2011. 44, 45, 46, 50, 54, 64
- [15] B. Bell and B. Cowie. The characteristics of formative assessment in science education. *Science education*, 85(5):536–553, 2001. 29
- [16] 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 *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual*, pages S3B–1. IEEE, 2007. 43, 44, 49, 63
- [17] A. Berns, A. Gonzalez-Pardo, and D. Camacho. Game-like language learning in 3-D virtual environments. *Computers & Education*, 60(1):210–220, 2013. 104
- [18] A. Berns, M. Palomo-Duarte, J. M. Dodero, and C. Valero-Franco. Using a 3D online game to assess students' foreign language acquisition and communicative competence. In *Scaling up learning for sustained impact*, pages 19–31. Springer, 2013. 104
- [19] A. Carreras Marín, Y. Blasco, M. Badia-Miró, M. Bosch Princep, 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. 39, 40, 41, 51, 54, 62
- [20] 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. 38, 40, 41, 50

BIBLIOGRAFÍA

- [21] 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. 45
- [22] M. A. Chatti, V. Lukarov, H. Thüs, A. Muslim, A. M. F. Yousef, U. Wahid, C. Greven, A. Chakrabarti, and U. Schroeder. Learning analytics: Challenges and future research directions. 2014. 5
- [23] 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, 16
- [24] C.-h. Chen. The implementation and evaluation of a mobile self-and peer-assessment system. *Computers & Education*, 55(1):229–236, 2010. 23
- [25] R. Cole, S. Purao, M. Rossi, and M. Sein. Being proactive: where action research meets design research. *ICIS 2005 Proceedings*, page 27, 2005. 75
- [26] 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
- [27] B. Communiqué. Making the most of our potential: Consolidating the european higher education area. In *2012 Ministerial Conference and 3rd Bologna Policy Forum Documents*, 2012. 3
- [28] 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
- [29] 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. 104

- [30] D. D. Curtis. The assessment of generic skills. *Generic skills in vocational education and training: Research findings*, pages 136–56, 2004. 1
- [31] B. De La Harpe, A. Radloff, and J. Wyber. Quality and generic (professional) skills. *Quality in Higher Education*, 6(3):231–243, 2000. 1
- [32] 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. 3
- [33] H. Dixon. What is teamwork? *The Canadian Journal of Medical Radiation Technology/CAMRT*, 23(2):79–80, 1992. 28
- [34] 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. 33, 54
- [35] M. P. Duarte, I. M. Buló, 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. 83
- [36] 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. 30
- [37] 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. 83
- [38] 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

BIBLIOGRAFÍA

- [39] 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
- [40] 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. 95, 104
- [41] 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. 38, 39, 40, 41, 50
- [42] 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. 45, 46, 48, 51, 54, 57, 64
- [43] A. Fisher. *Critical thinking: An introduction*. Cambridge University Press, 2011. 28
- [44] 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
- [45] M. Fowler. *Domain-specific languages*. Pearson Education, 2010. 79
- [46] 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. 98
- [47] S. M. Gass. *Second language acquisition: An introductory course*. Routledge, 2013. 28

- [48] F. Gates. *Teachers Know Best. Making Data Work For Teachers and Students*. Bill & Melinda Gates Foundation, 2015. 16
- [49] J. González and R. Wagenaar. Tuning educational structures in europe ii. universitites’ contribution to the bologna process. *University of Deusto*, 2005. 3, 27, 28
- [50] 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. 44, 45, 46, 51, 64
- [51] 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. 52
- [52] 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. 29
- [53] A. R. Hevner. Interview with Alan R. Hevner on “Design Science”. *Business & Information Systems Engineering (BISE)*, 1(1):126–129, 2009. 6
- [54] 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. 21
- [55] 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. 104
- [56] G. Hughes. Towards a personal best: a case for introducing ipsative assessment in higher education. *Studies in Higher Education*, 36(3):353–367, 2011. 30
- [57] A. Huhta. Diagnostic and formative assessment. *The handbook of educational linguistics*, 33:469–483, 2008. 29

BIBLIOGRAFÍA

- [58] J. L. C. Izquierdo and J. Cabot. Community-driven language development. In *Modeling in Software Engineering (MISE), 2012 ICSE Workshop on*, pages 29–35. IEEE, 2012. 74
- [59] B. Kitchenham, R. Pretorius, D. Budgen, O. Pearl Brereton, M. Turner, M. Niazi, and S. Linkman. Systematic literature reviews in software engineering - a tertiary study. *Information and Software Technology*, 52(8):792–805, Aug. 2010. 17
- [60] R. Lacuesta, G. Palacios, and L. Fernández. Active learning through problem based learning methodology in engineering education. In *Frontiers in Education Conference, 2009. FIE'09. 39th IEEE*, pages 1–6. IEEE, 2009. 32, 42, 43, 44, 49, 54, 63
- [61] 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. 38, 39, 40, 41, 43, 50, 62
- [62] D. Laurillard. Teaching as a design science. *Building pedagogical patterns for learning and technology*, 2012. 5, 7
- [63] C. S. E. Lidz. *Dynamic assessment: An interactional approach to evaluating learning potential*. Guilford Press, 1987. 30
- [64] H. Malehorn. Ten measures better than grading. *The Clearing House*, 67(6):323–324, 1994. 37, 62
- [65] 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. 43, 49
- [66] J. E. P. Martinez, J. García Martín, and A. S. Alonso. Teamwork competence and academic motivation in computer science engineering studies. In *Global Engineering Education Conference (EDUCON), 2014 IEEE*, pages 778–783. IEEE, 2014. 38, 40, 41, 51

- [67] 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 *Global Engineering Education Conference (EDUCON), 2013 IEEE*, pages 436–441. IEEE, 2013. 39, 40, 41, 51
- [68] 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
- [69] V. Midoro. Guidelines on adaptation of the UNESCO ICT competency framework for teachers. 2013. 4
- [70] P. Moreno-Ger, I. Martinez-Ortiz, M. Freire, B. Manero, and B. Fernandez-Manjon. Serious games: A journey from research to application. In *Frontiers in Education Conference (FIE), 2014 IEEE*, pages 1–4. IEEE, 2014. 104
- [71] 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. Accessed: 2016-05-01. 32
- [72] B. J. Oates. *Researching information systems and computing*. Sage, 2005. 13, 112
- [73] B. Oliver. Graduate attributes as a focus for institution-wide curriculum renewal: innovations and challenges. *Higher Education Research & Development*, 32(3):450–463, 2013. 40, 41, 50
- [74] J. Ortega Valiente and J. Reinoso Peinado. New educational approach based on the use of wiki platforms in university environments. In *Next Generation Web Services Practices (NWeSP), 2011 7th International Conference on*, pages 280–284. IEEE, 2011. 83

BIBLIOGRAFÍA

- [75] 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. 110, 132
- [76] 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. 83
- [77] 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. British Computer Society, 2008. 17, 21
- [78] 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 *Education Engineering (EDUCON), 2010 IEEE*, pages 1511–1516. IEEE, 2010. 40, 41, 50
- [79] 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. 16
- [80] QAA. *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education*. The Quality Assurance Agency for Higher Education, 2006. 30
- [81] 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)*, number 5, pages 377–383. World Scientific and Engineering Academy and Society (WSEAS), 2008. 43, 50

- [82] A. Rayon Jerez, M. Guenaga, and A. Núñez. A web platform for the assessment of competences in mobile learning contexts. In *Global Engineering Education Conference (EDUCON), 2014 IEEE*, pages 321–329. IEEE, 2014. 45, 46, 47, 49, 54, 57, 64
- [83] C. Redecker and [U+FFFD] Johannessen. Changing Assessment — Towards a New Assessment Paradigm Using ICT. *European Journal of Education*, 48(1):79–96, 2013. 25
- [84] P. Reimann. Design-based research. In *Methodological Choice and Design*, pages 37–50. Springer, 2011. 7
- [85] A. J. Reinoso, M. Palomo-Duarte, J. T. Tocino-Garcia, J. Ortega-Valiente, and A. Balderas. Evaluación del trabajo individual y grupal en un wiki. In *III Congreso Internacional Sobre Aprendizaje, Innovación y Competitividad. CINAIC 2015*. 133
- [86] 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. 40, 41, 51
- [87] S. Rodriguez-Donaire, B. A. García, and S. O. Del Olmo. e-portfolio: a tool to assess university students’skills. In *Information Technology Based Higher Education and Training (ITHET), 2010 9th International Conference on*, pages 114–124. IEEE, 2010. 43, 50
- [88] A. Rojas-Vargas. *La prueba escrita*. Ministerio de Educación Pública de Costa Rica. Dirección de desarrollo curricular. Departamento de Evaluación de los Aprendizajes, 2008. 31
- [89] 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. 40, 41, 51

BIBLIOGRAFÍA

- [90] 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. 43, 44, 50, 63
- [91] 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. 39, 40, 41, 49, 62
- [92] G. Siemens and R. S. d Baker. Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pages 252–254. ACM, 2012. 45
- [93] A. I. Starcic. Sustaining teacher’s professional development and training through web-based communities of practice. In *Applications and the Internet, 2008. SAINT 2008. International Symposium on*, pages 317–320. IEEE, 2008. 38, 42, 43, 51
- [94] 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, 49, 52
- [95] 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. 32
- [96] G. Trentin. Using a wiki to evaluate individual contribution to a collaborative learning project. *Journal of Computer Assisted Learning*, 25(1):43–55, 2009. 83
- [97] 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. 28

- [98] R. W. Tyler, R. M. Gagné, M. Scriven, et al. *Perspectives of curriculum evaluation*, volume 1. Rand McNally Chicago, 1967. 29
- [99] 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 *Frontiers in Education Conference, 2013 IEEE*, pages 999–1003. IEEE, 2013. 43, 49
- [100] 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. 32, 42, 43, 50
- [101] 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. 21
- [102] J. S. Wiggins. *Paradigms of personality assessment*. Guilford Press, 2003. 33
- [103] F. Yang, F. W. Li, and R. W. Lau. A fine-grained outcome-based learning path model. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, 44(2):235–245, 2014. 42, 43, 49
- [104] 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

Declaración

I herewith declare that I have produced this work without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This work has not previously been presented in identical or similar form to any examination board.

The dissertation work was conducted from 20XX to 2016 under the supervision of Name Surname and Name Surname at the University Cádiz.

Puerto Real,

This dissertation was finished writing in Jerez de la Frontera on 21 de abril
de 2016

This page is intentionally left blank