



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, enero 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, enero 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

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.

Acknowledgements

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 tristisque. Aenean fermentum, dolor eget tincidunt faucibus, risus lorem feugiat elit, sagittis malesuada eros ligula in odio. Pellentesque ac libero lobortis justo bibendum laoreet. Cras egetas 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 egetas 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 egetas. 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.

Donec nulla enim, scelerisque quis dignissim ut, vehicula sit amet felis. Ut tristisque pulvinar aliquet. Proin vitae odio nibh. Sed est dolor, malesuada et lobortis vitae, tempor quis tellus. Proin mollis lacus eget arcu tempor vitae bibendum nisi adipiscing. In sollicitudin pretium dapibus. Sed accumsan imperdiet diam quis pellentesque.

Sed in lacinia lectus. Nullam ac magna in erat blandit posuere nec nec ante. Pellentesque habitant morbi tristisque senectus et netus et malesuada fames ac turpis egetas. Maecenas ipsum neque, fringilla sed sodales id, scelerisque et erat. Morbi placerat mauris vel elit ultricies sed molestie ipsum faucibus. Nunc vel sem orci, ac placerat justo. Nulla varius nisi mauris. Vivamus sed ligula felis, ac mattis purus. Vivamus

felis nibh, bibendum et aliquam sed, feugiat a elit. Morbi imperdiet libero vitae lacus viverra tempus. Praesent ultricies fermentum urna eget varius. Nunc blandit, augue eget vehicula venenatis, ligula eros posuere tellus, non sollicitudin tortor velit quis augue. Nulla rhoncus fringilla dolor, vel cursus ligula varius id. Vestibulum et eros eros, ut euismod elit. Praesent placerat libero sed lorem mollis sed cursus erat laoreet.

Suspendisse in fermentum nulla. Donec blandit ultricies felis, non suscipit turpis sagittis eget. Donec eget enim in erat auctor lobortis eu sagittis mauris. Aliquam vitae nibh nisi. Integer nulla erat, feugiat quis vulputate vel, vulputate id erat. Morbi dolor dui, volutpat congue fermentum tempus, rutrum et quam. Vivamus aliquam, sapien in ultrices dictum, tortor tellus dictum justo, eu viverra velit tellus congue nibh. Suspendisse potenti. Nullam at massa id leo sollicitudin ultricies ut et nibh. Vivamus fermentum pellentesque ante et lobortis. Etiam eros justo, rutrum at varius et, porttitor ac lorem. Aenean viverra, mauris eget imperdiet aliquet, risus quam tincidunt risus, quis elementum enim diam a ipsum. Ut et arcu nunc, a viverra mi. Nullam egestas lectus vel ipsum sodales eget consectetur metus facilisis. Proin egestas magna sed felis mattis ultrices.

Morbi a lectus vitae lacus malesuada tincidunt. Vestibulum eu quam justo. Integer hendrerit posuere augue, in fringilla nulla feugiat ut. Fusce vel libero sed tellus posuere molestie in sed justo. Nunc orci magna, aliquam id dictum sit amet, bibendum vestibulum augue. Maecenas luctus ultricies elit, et aliquet justo rutrum sed. Proin mauris dui, cursus quis tempus non, rhoncus nec ante. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Fusce ac nunc non felis fermentum ultrices. Curabitur nibh felis, convallis id consequat rhoncus, elementum sodales ipsum. In hac habitasse platea dictumst. In quis nulla eu sapien mattis cursus nec semper leo. Maecenas sagittis viverra quam, vel sodales elit placerat et. Sed condi-

mentum ultricies mattis. Pellentesque scelerisque convallis lobortis. In
tempus lorem nec lorem blandit sodales.

Gracias,

Antonio

enero 2016

Índice general

| | |
|--|-------------|
| Índice de figuras | xv |
| Índice de tablas | xvii |
| 1 Introducción | 1 |
| 1.1 Motivación | 3 |
| 1.2 Contexto | 7 |
| 1.3 Objetivos y preguntas de investigación | 10 |
| 1.4 Estrategia de investigación | 12 |
| 1.4.1 Diseño y creación | 12 |
| 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 |
| 2.3.1 Localización de la literatura | 24 |
| 2.3.2 Extracción de los datos | 25 |
| 2.3.2.1 Competencias genéricas | 27 |

ÍNDICE GENERAL

| | | |
|-----------|--|-----------|
| 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 | 70 |
| 4.3.1 | Características | 71 |
| 4.3.2 | Requisitos | 72 |
| 4.3.3 | Herramienta | 75 |
| 5 | Evaluación | 77 |
| 5.1 | Introducción | 78 |
| 5.1.1 | Entornos de aplicación | 78 |
| 5.1.1.1 | Wikis | 78 |
| 5.1.1.1.1 | AssessMediaWiki (AMW) | 79 |
| 5.1.1.1.2 | Método | 81 |
| 5.1.1.1.3 | Evaluación del trabajo en el wiki | 81 |

ÍNDICE GENERAL

| | | |
|-----------|--|-----|
| 5.1.1.1.4 | Evaluación de competencias genéricas | 86 |
| 5.1.1.1.5 | Indicadores | 87 |
| 5.1.1.2 | VLE | 90 |
| 5.1.1.2.1 | SASQL y EvalCourse | 90 |
| 5.1.1.2.2 | Método | 93 |
| 5.1.1.2.3 | Indicadores | 95 |
| 5.1.1.3 | Mundos virtuales | 97 |
| 5.1.1.3.1 | VWQL y EvalSim | 98 |
| 5.1.1.3.2 | Método | 99 |
| 5.1.1.3.3 | Indicadores | 101 |
| 5.1.2 | Publicaciones | 103 |
| 5.1.3 | Cuestionarios y entrevistas | 103 |
| 5.2 | AMW aplicado a MediaWiki | 103 |
| 5.2.1 | Evaluación | 104 |
| 5.2.1.1 | Indicadores cuantitativos | 104 |
| 5.2.1.2 | Indicadores cualitativos | 105 |
| 5.2.1.3 | Resumen | 105 |
| 5.2.2 | Resultados | 107 |
| 5.2.2.1 | Indicador del trabajo en equipo (M_1) | 109 |
| 5.2.2.2 | Indicador de las habilidades de comunicación y la aplicación del conocimiento (M_2) | 110 |
| 5.2.2.3 | Indicador de habilidades individuales (M_3) | 111 |
| 5.2.2.4 | Indicador del pensamiento crítico (M_4) | 113 |
| 5.2.2.5 | Entregable final (M_5) | 114 |
| 5.2.3 | Discusión | 116 |
| 5.3 | EvalCourse aplicado a entornos de aprendizaje virtual | 117 |
| 5.3.1 | Evaluación | 117 |
| 5.3.1.1 | Extracción de indicadores del foro | 118 |
| 5.3.1.2 | Extracción de indicadores del taller | 121 |
| 5.3.2 | Resultados | 124 |
| 5.4 | EvalSim aplicado a los mundos virtuales | 125 |
| 5.5 | Evaluación | 125 |
| 5.6 | Conclusiones | 125 |

ÍNDICE GENERAL

| | |
|--|------------|
| 6 Conclusiones y trabajo futuro | 127 |
| Bibliografía | 129 |

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 |
| 5.1 | Rúbrica de AMW | 80 |
| 5.2 | Ejemplo de retroalimentación formativa y la contribución de wiki evaluada | 81 |
| 5.3 | Ejemplo de flujo de trabajo para la evaluación cualitativa del wiki utilizando AMW | 82 |
| 5.4 | Ciclo de contraste de hipótesis para la evaluación en wikis | 87 |
| 5.5 | Metamodelo de SASQL | 91 |
| 5.6 | Ciclo de contraste de hipótesis utilizando EvalCourse | 94 |
| 5.7 | Interacción en el foro en un periodo de tiempo. | 96 |
| 5.8 | Ciclo de contraste de hipótesis con EvalSim | 99 |
| 5.9 | Distribución del trabajo de los miembros del proyecto P_{13} | 111 |
| 5.10 | Resumen del desempeño de los estudiantes en el foro en las dos competencias. | 119 |
| 5.11 | Graph of students' interaction in forum case study. | 120 |
| 5.12 | Grafo que muestra las diferentes calificaciones que se autoasignaron los estudiantes y la que les dieron sus compañeros. | 122 |

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

| | | |
|------|---|-----|
| 5.1 | Resumen de las competencias evaluadas para cada tipo de indicador | 89 |
| 5.2 | Información sobre la participación en el foro de los estudiantes en un periodo concreto de tiempo | 95 |
| 5.3 | Información sobre las respuestas de una sola palabra dadas por los estudiantes en el role-play | 101 |
| 5.4 | Información sobre las palabras por turnos utilizadas por los estudiantes en el role-play | 102 |
| 5.5 | Resumen de indicadores para cada aproximación. | 106 |
| 5.6 | Grupos de ejemplo y sus miembros. | 108 |
| 5.7 | Grading of collaborative work (M_1) in the P_{13} project. | 110 |
| 5.8 | (M_2) para el proyecto P_7 | 112 |
| 5.9 | M_2 para el proyecto P_4 | 112 |
| 5.10 | (M_3) para los miembros de P_4 y P_{13} | 113 |
| 5.11 | (M_4) para los miembros del proyecto P_{13} | 114 |
| 5.12 | Calificaciones dadas por el profesor al proyecto P_{13} | 115 |
| 5.13 | Resumen de calificaciones finales de los estudiantes | 117 |
| 5.14 | Listado de intervenciones de los estudiante en el foro. | 119 |
| 5.15 | Desempeño de los estudiantes en la competencia de la autocrítica. . | 123 |

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 [35, 65]. La investigación educativa muestra que dichas competencias se enseñan con mayor eficacia en el periodo formativo del individuo [29].

La evaluación de competencias genéricas se ha abordado desde diferentes perspectivas. En la revisión de la literatura realizada por Curtis [28] 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 [25], 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 [30]. 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 [101].

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* [46], apoyado por el Lifelong Learning Programme de la Unión Europea [36], 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 [90]. 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 [66], 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 [22, 41] 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 [26]. 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 [21]. 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* [59].

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 [24].

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 [50]

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 [59]. 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 [81]. 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 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

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

1.3 Objetivos y preguntas de investigación

genéricas.

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. INTRODUCCIÓN

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 [69]. 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

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 [76]. 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 [22]. Según el informe *Making Data Work For Teachers and Students*, de la *Fundación Bill & Melinda Gates* [45], 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 [55]. 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 [74]. 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 [98] y Hevner [51]. Usamos el primero, ya que es el recomendado en el SMS descrito por Petersen [74].

- 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 [23].
- 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*) [80], 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* [46]. 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 [46]. |
| 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 [93]. |
| Comunicación | Habilidad para comunicarse de manera tanto oral como escrita en la lengua materna [46]. |
| Creatividad | Capacidad para crear nuevas ideas [46]. |
| Cultural | Aprecio y respeto por la diversidad y la multiculturalidad [46]. |
| Emprendimiento | Capacidad para tomar la iniciativa y espíritu de empresa [46]. |
| Gestión de proyectos | Habilidades para diseñar y gestionar proyectos [46]. |
| Habilidades interpersonales | Capacidad de la persona para comunicarse e interactuar con otras personas [46]. |
| Investigación | Capacidad para llevar a cabo la investigación en un nivel apropiado [46]. |
| Liderazgo | Habilidad para motivar a la gente y conducirlos hacia un objetivo común [46]. |
| Pensamiento crítico | Habilidad para interpretar, analizar y evaluar ideas y argumentos [40]. |
| Planificación y gestión del tiempo | Capacidad de planificar y gestionar el tiempo de manera efectiva [46]. |
| Resolución de problemas | Habilidad para identificar, plantear y resolver problemas [46]. |
| Responsabilidad | Capacidad para actuar con responsabilidad social y conciencia cívica [46]. |
| Segundo idioma | Capacidad de los estudiantes para comunicar sus ideas en un segundo idioma [44]. |
| TIC | Habilidades en el uso de las tecnologías de la información y de la comunicación [46]. |
| Toma de decisiones | Capacidad para tomar decisiones razonadas [46]. |
| Trabajo autónomo | Capacidad para trabajar de forma autónoma [46]. |
| Trabajo en equipo | Trabajo realizado por un conjunto de personas en aras de un objetivo común [31]. |

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 [14]. 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 [94].
- 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 [49].
- 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 [54].
- 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 [60].

- **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 [77].
- **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 [33].
- **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 [53].

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 [84]. 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 [91].
- 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 [97] 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 [68].
- La *herramienta de seguimiento* es una herramienta para monitorizar el trabajo del estudiante a lo largo del semestre. En el trabajo presentado en [56] 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 [99]. 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 [32]. 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.

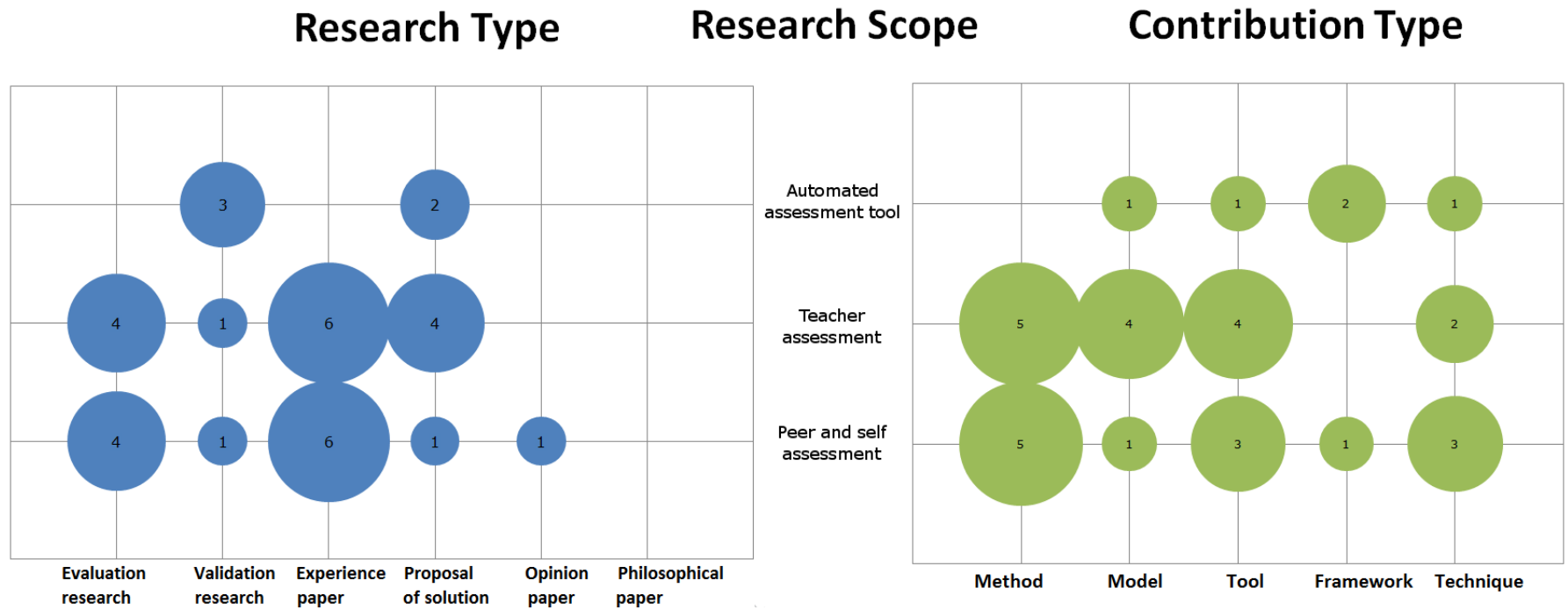


Figura 2.2: Ámbito de trabajos distribuidos según tipo de investigación y según tipo de contribución.

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 [61].

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 [89]. 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 [63]. En [58] 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 [38] 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 [19] 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 [64].

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 [38]. 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 [58, 87]. Además, se puede dar el caso que la autoevaluación no se ajuste del todo a la realidad del desempeño del estudiante. Por ejemplo, en [18] hay notables

2. ESTADO DEL ARTE

| COMPETENCIA | CANT. | TRABAJOS |
|------------------------------------|-------|------------------------------|
| Análisis | 2 | [18, 58] |
| Aprendizaje permanente | 1 | [70] |
| Comunicación | 5 | [18, 63, 64, 70, 85] |
| Creatividad | 1 | [75] |
| Cultural | 1 | [70] |
| Emprendimiento | 2 | [1, 19] |
| Gestión de proyectos | 1 | [63] |
| Habilidades interpersonales | 1 | [63] |
| Investigación | 1 | [70] |
| Liderazgo | 1 | [63] |
| Pensamiento crítico | 1 | [6] |
| Planificación y gestión del tiempo | 1 | [63] |
| Resolución de problemas | 1 | [70] |
| Responsabilidad | 2 | [18, 85] |
| Segundo idioma | 3 | [64, 82, 87] |
| TIC | 3 | [58, 64, 70] |
| Trabajo autónomo | 1 | [58] |
| Trabajo en equipo | 6 | [18, 38, 58, 63, 70, 75, 85] |

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 | - | [82] |
| | Observación sistemática | Rúbrica | [70] |
| Formativa | Pruebas escritas | Cuestionario | [1] |
| | Observación sistemática | Rúbrica | [6, 75] |
| Sumativa | Pruebas orales | | [18, 58] |
| | Pruebas escritas | Cuestionario | [64] |
| | | | [19, 38, 63, 85, 87] |

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 [89], metodologías PBL [56] o experiencias basadas en videoconferencias [97].

En [100] 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

La 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 | [78] |
| Comunicación | 6 | [15, 56, 62, 78, 83, 100] |
| Emprendimiento | 2 | [78, 97] |
| Pensamiento crítico | 2 | [7, 56] |
| Planificación y gestión del tiempo | 1 | [56] |
| Resolución de problemas | 5 | [7, 15, 62, 83, 96] |
| Trabajo autónomo | 1 | [58] |
| Trabajo en equipo | 5 | [15, 56, 62, 78, 83] |

Tabla 2.9: Competencias evaluadas directamente por el profesor

| MÉTODOS DE EVALUACIÓN | TÉCNICAS DE EVALUACIÓN | INSTRUMENTOS DE EVALUACIÓN | TRABAJO |
|----------------------------|----------------------------|----------------------------|-------------|
| Auténtica | Actividades sin determinar | - | [89] |
| Con referencia al criterio | | | [86] |
| Formativa | Pruebas orales | Entrevista | [97] |
| | Observación sistemática | Herramienta de seguimiento | [15, 56] |
| | | Rúbrica | [83] |
| Sumativa | Actividades sin determinar | - | [7, 62, 78] |
| | | | [100] |
| | Pruebas escritas | Cuestionario | [96] |

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 [86] 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 [56]. 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 [15], 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) [13, 47]. 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 [20].

Y finalmente encontramos trabajos que se basan en técnicas de *learning analytics* [39, 79], 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 [88].

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 [47] 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 [13] 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, 13, 79] |
| Creatividad | 1 | [4] |
| Emprendimiento | 1 | [47] |
| Gestión de proyectos | 1 | [4] |
| Habilidades interpersonales | 2 | [4, 79] |
| Investigación | 1 | [4] |
| Liderazgo | 1 | [4] |
| Pensamiento crítico | 1 | [4] |
| Planificación y gestión del tiempo | 1 | [4] |
| Resolución de problemas | 1 | [47] |
| Responsabilidad | 1 | [4] |
| Toma de decisiones | 1 | [4] |
| Trabajo autónomo | 1 | [4] |
| Trabajo en equipo | 2 | [4, 39] |

Tabla 2.11: Competencias evaluadas de forma semiautomática

| MÉTODOS DE EVALUACIÓN | TÉCNICAS DE EVALUACIÓN | INSTRUMENTOS DE EVALUACIÓN | TRABAJOS |
|-----------------------|-------------------------------|----------------------------|-------------------------------|
| Auténtico | Indicadores basados en logros | Juegos serios | [13] |
| Formativo | | | Indicadores de trabajo en VLE |
| | [39, 79] | | |
| 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 [79]. 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 [39]. 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 |
|-------|---|-----------------------|----------------------|---|
| [100] | A Fine-Grained Outcome-Based Learning Path Model | proposal of solution | model | Teacher assessment |
| [79] | A web platform for the assessment of competences in Mobile Learning Contexts | validation research | framework | Automated assessment tool |
| [62] | Acquired Skills With The Implementation Of New Evaluation Methods At University Rey Juan Carlos | experience paper | model | Teacher assessment |
| [56] | Active learning through problem based learning methodology in engineering education | experience paper | method | Teacher assessment |
| [15] | 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 |
| [87] | 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 |
| [96] | 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 |
| [90] | Criteria and standards of generic competences at bachelor degree level: A review study | evaluation research | method | Review study |

| | | | | |
|------|--|----------------------|-----------|---|
| [97] | Developing entrepreneurial accounting and finance competency using the ELLEIEC Virtual Centre for Enterprise | experience paper | tool | Teacher assessment |
| [38] | e-Learning and Team-based Learning. Practical Experience in Virtual Teams | experience paper | framework | Peer and self-assessment |
| [78] | Engineering Students Performance Evaluation of Generic Skills Measurement: ESPEGS Model | validation research | model | Teacher assessment |
| [83] | 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 |
| [13] | From Behavioral Indicators to Contextualized Competence Assessment | proposal of solution | framework | Automated assessment tool |
| [70] | Graduate attributes as a focus for institution-wide curriculum renewal: innovations and challenges | opinion paper | model | Peer and self-assessment |
| [86] | Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences | proposal of solution | tool | Teacher assessment |
| [19] | International creative tension study of university students in South Korea and Finland | evaluation research | tool | Peer and self-assessment |
| [75] | Measuring collaboration and creativity skills through rubrics: Experience from UTPL collaborative social networks course | evaluation research | method | Peer and self-assessment / Teacher assessment |
| [58] | 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 |
| [64] | Self-video recording for the integration and assessment of generic competencies | experience paper | technique | Peer and self-assessment |
| [47] | Serious Games for the Development of Employment Oriented Competences | validation research | tool | Automated assessment tool |
| [85] | Soft Skills: A Comparative Analysis Between On-line and Classroom Teaching | experience paper | method | Peer and self-assessment |
| [89] | Sustaining Teacher's Professional Development and Training through Web-Based Communities of Practice | evaluation research | tool | Teacher assessment |
| [82] | Teaching And Learning Through Projects Using The ICT: Practice Of The English Writing Through Business Documents | experience paper | method | Peer and self-assessment |
| [63] | Teamwork competence and academic motivation in computer science engineering studies | validation research | method | Peer and self-assessment |
| [18] | The promotion and assessment of generic skills from interdisciplinary teaching teams | experience paper | method | Peer and self-assessment / Teacher assessment |
| [39] | 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 [90], 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 [48].

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 [56]. 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 [18].

Dentro de la evaluación semiautomática se seleccionaron 5 trabajos, entre los que se encuentran algunos basados en juegos serios [13, 32] y otros basados en el análisis de los procesos de aprendizaje [39, 79].

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

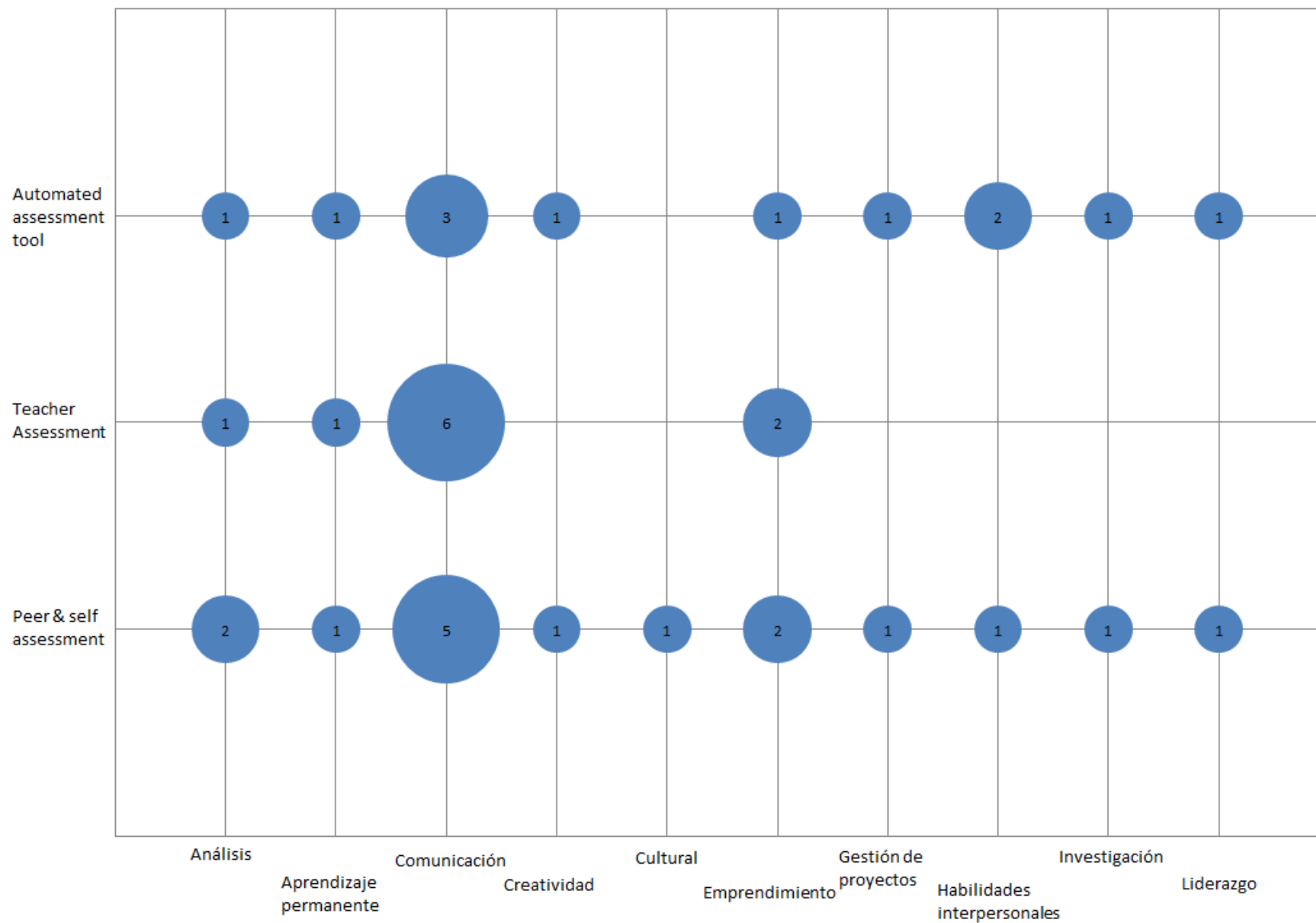


Figura 2.3: Competencias genéricas evaluadas con cada método (1 de 2).

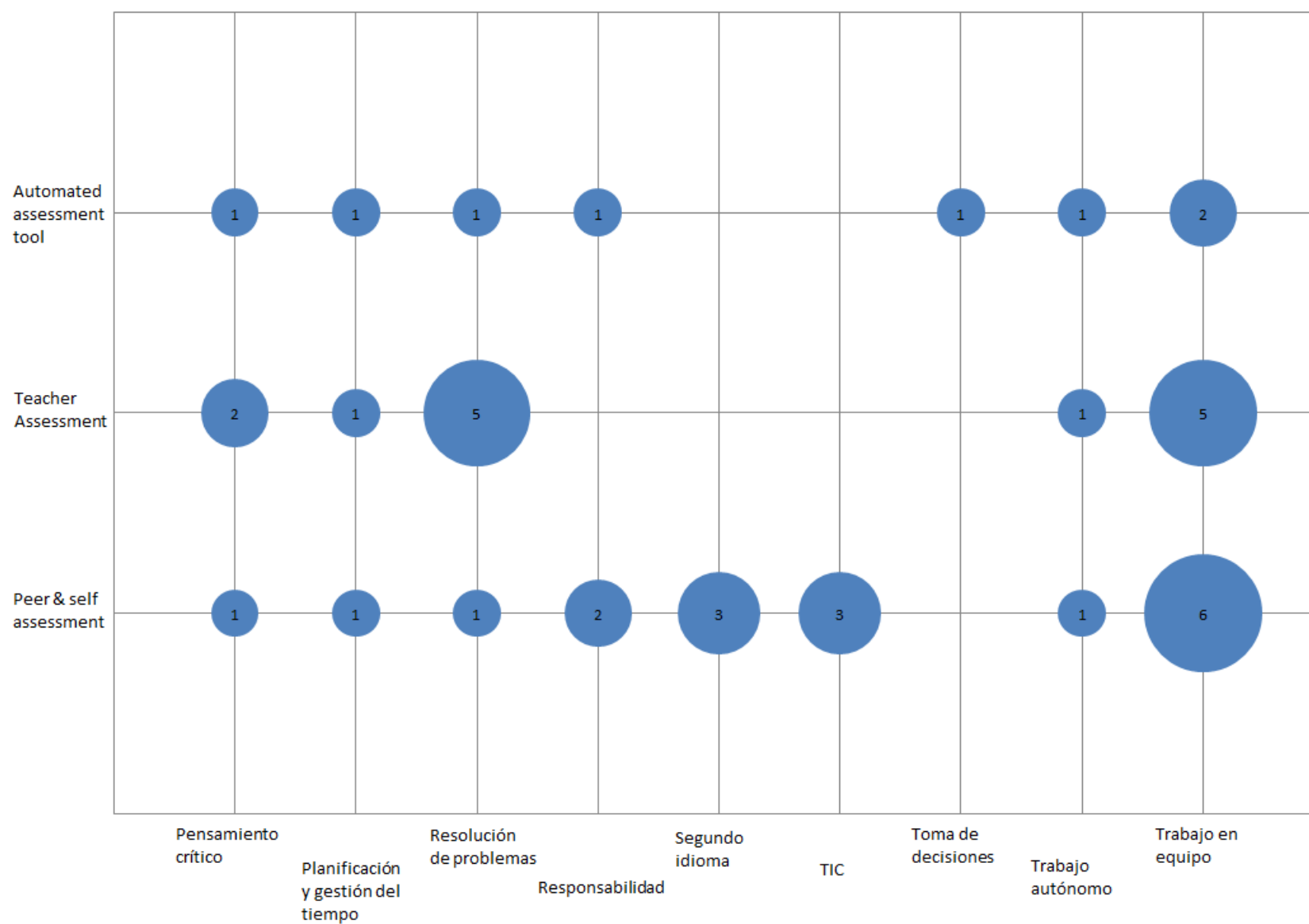


Figura 2.4: Competencias genéricas evaluadas con cada método (2 de 2).

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> | [79] | | |
| <i>Comunicación</i> | | | |
| <i>Habilidades interpersonales</i> | | | |
| <i>Planificación</i> | | | |
| <i>Trabajo en equipo</i> | [39] | | |

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 [39, 79].

En [79] 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 [39] 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. Una vez trabajadas estas competencias y finalizados estos 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 puede no ser válido 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 [61]. 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 [18].

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 [58, 87]. 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 [56, 86] e incluso uno de ellos descarta repetir la experiencia en los siguientes cursos debido al sobreesfuerzo que le supuso [15]. 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 [13, 47]. 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 [39, 79]. 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

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 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. Estas evaluaciones podrán ser refinadas por el propio profesor hasta llegar a unas evidencias que le sean válidas. También podrán ser utilizadas y modificadas por

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

otros profesores que busquen adaptarlo a su contexto local o a las competencias que ellos quieren realizar.

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 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, ya sea porque tienen un examen o porque tienen que subir una tarea. Todas las acciones de los estudiantes quedan almacenadas en el registro de las actividades de aprendizaje, y estos registros podrían ser analizados para comprender el proceso de aprendizaje que se está desarrollando mediante *learning analytics*, contexto en el que sitúa el método DBA.

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, escribiendo y respondiendo mensajes en los foros o enviando actividades, como la participación pasiva, cuando simplemente se lee una página, se leen los mensajes del foro o se descargan los apuntes.

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 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

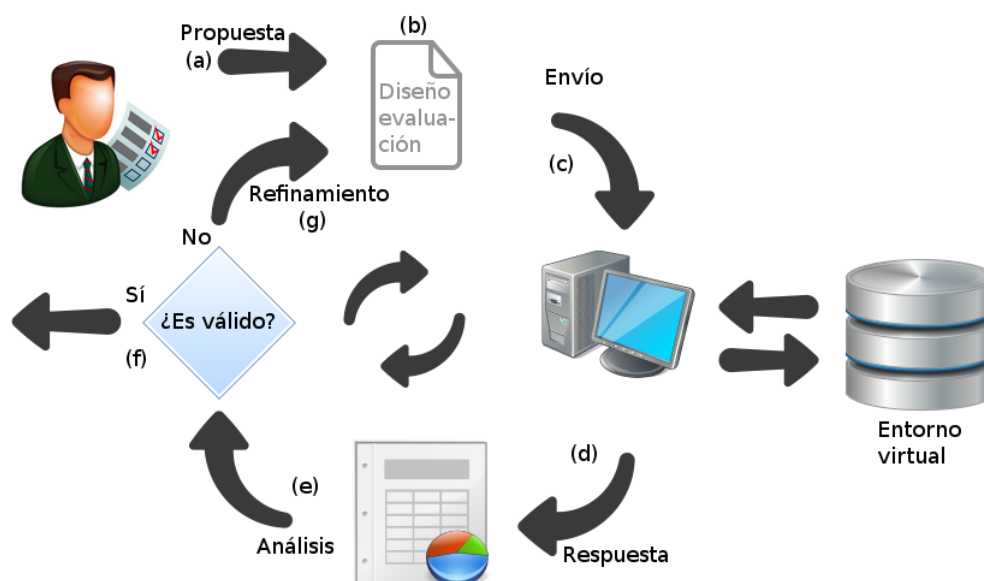


Figura 4.1: Diagrama del ciclo de contraste de hipótesis

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 virtual para la evaluación de alguna competencia genérica (a). Por ejemplo,

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

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 5.1.1.
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).

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 puede utilizar en contextos educativos reales. De hecho, el método ya ha sido utilizado en diferentes asignaturas de la Universidad de Cádiz y los resultados de la investigación han sido publicados en revistas y congresos del área de las TEL [8, 9, 10, 11, 12].
- *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.
- *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 método DBA permite la colaboración entre los investigadores, personal con perfil tecnológico del área de la informática, y profesores sin perfil tecnológico de cualquier

4. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

área. Para ello, se implementa una herramienta que permita aplicar el método a profesores sin conocimientos técnicos, y así tanto investigadores como profesores pueden poner en común su experiencia y los resultados que van obteniendo.

- *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.
- *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.
- *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 sé. 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 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 del indicador.

2. Evaluación escalable

El método para la evaluación de competencias genéricas es escalable y no supone al profesor un esfuerzo que éste no pueda abordar. El método se implementa en una herramienta que se alinea con las actividades aprendizaje. El profesor puede consultar los indicadores del registro con una simple petición a la herramienta. Entonces, esta procesa la petición y la información es devuelta en formatos que el profesor puede 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. MÉTODO PARA LA EVALUACIÓN DE COMPETENCIAS GENÉRICAS

4. Accesibilidad

No es necesario 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 implementa el método es usable y sencilla para que los profesores puedan utilizarla sin requerirles conocimientos técnicos, y los formatos a los que se exporta la información son 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 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.

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 [42]. 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

Evaluación

En este capítulo se evalúa la idoneidad del método propuesto para la evaluación de competencias genéricas. Este capítulo se estructura de la siguiente manera:

- En primer lugar se describen los entornos de aplicación y las características de los procedimientos que se han utilizado para llevar a cabo la evaluación:
 - Publicaciones
 - Cuestionarios
- En segundo lugar se muestran los resultados de la aplicación en cada una de las actividades de aprendizaje para los que se ha implementado el método:
 - Wikis: AMW
 - VLE: SASQL Y EvalCourse
 - Mundos virtuales: VWQL Y EvalSim
- Finalmente se presentan las conclusiones de estas evaluaciones, tanto de los cuestionarios como de las publicaciones.

5. EVALUACIÓN

5.1 Introducción

La evaluación del método se ha realizado desde dos perspectivas. Por un lado, cada una de las implementaciones realizadas se ha tratado de mostrar tanto en congresos y revistas de relevancia en el área de las TEL. Mientras que por otro lado, se han realizado cuestionarios y entrevistas a profesionales de la enseñanza para evaluar la idoneidad de los indicadores obtenidos. A continuación se introducen ambos enfoques.

5.1.1 Entornos de aplicación

Hay diversos entornos en los que se pueden desarrollar actividades de aprendizaje. Para la implementación del método se seleccionaron tres entornos diferentes: el primer entorno fue un wiki, entorno de trabajo online donde los usuarios crean y editan el contenido de forma colaborativa; el segundo fue un VLE, entorno de aprendizaje virtual donde se alojan cursos virtuales constituidos por una miríada de herramientas; y el tercero un mundo virtual, donde los estudiantes afrontan situaciones de la vida real en un entorno de simulación virtual.

En las próximas secciones se describirán cada una de las herramientas implementadas para aplicar el método en cada uno de los entornos virtuales: AssessMediaWiki para wikis, EvalCourse para VLEs y EvalSim para mundos virtuales.

5.1.1.1 Wikis

El uso educacional de los wikis para las experiencias de trabajo colaborativo está en auge debido a las numerosas ventajas que aporta sobre los modelos tradicionales [34]. 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 profesor 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 [92]. 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 [71]. 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 trabajo anterior se utilizó *StatMediaWiki* (SMW) ¹, una herramienta que proporciona al profesor 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. A partir de esa información cuantitativa se evaluó la distribución del trabajo, el trabajo en equipo y el liderazgo. Sin embargo, el aspecto cualitativo quedó fuera, ya que el experimento sólo consideró el número, el momento y el tamaño de las contribuciones [72].

Para completar el análisis cuantitativo proporcionado por SMW con un análisis cualitativo se desarrolló *AssessMediaWiki* (AMW) ². 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 profesor.

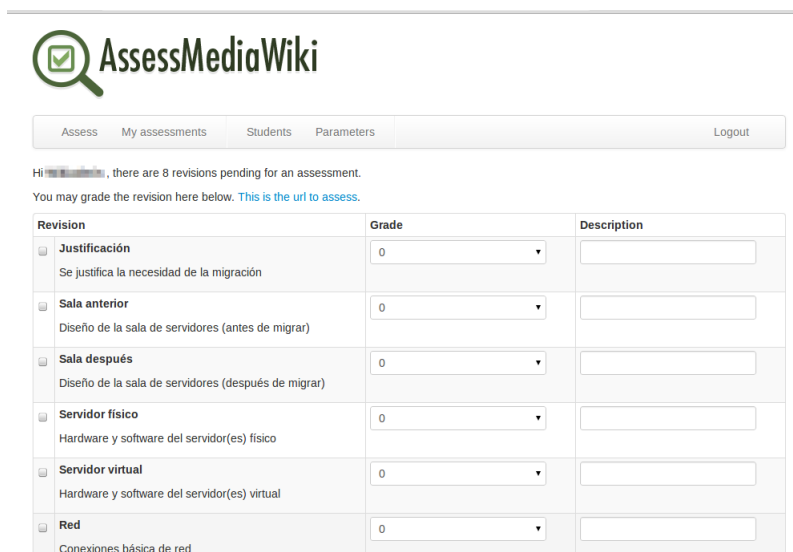
5.1.1.1 AssessMediaWiki (AMW)

La aplicación creada para poner en práctica este método es 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 profesor, a la vez que mantiene información sobre esas evaluaciones. AMW pone a disposición de los estudiantes una rúbrica previamente definida por el profesor para que realicen la evaluación (figura5.1).

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

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

5. EVALUACIÓN



The screenshot shows the AssessMediaWiki interface. 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 with a checkbox, a title, a description, a grade dropdown menu (all set to 0), and a description input field.

| 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

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 alumnos vayan haciendo. AMW implementa un 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 utilizando para ello una réplica similar a la que se utilizó en su evaluación, indicando las calificaciones que considera que merece y sus correspondientes justificaciones. Después el profesor revisará la disputa y pondrá la nota definitiva.

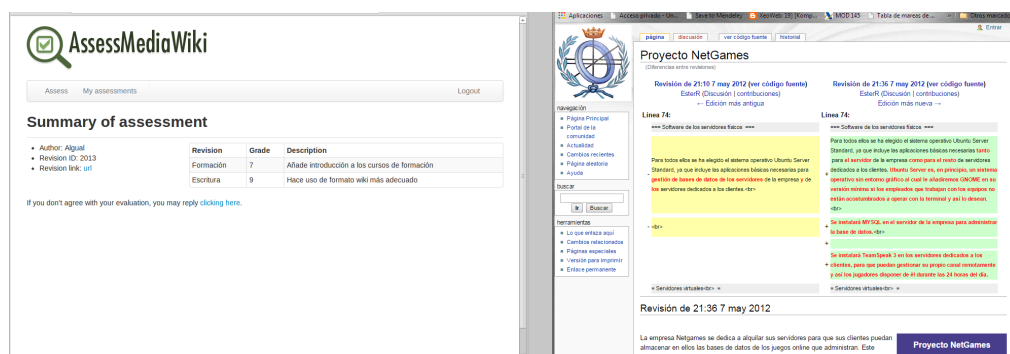


Figura 5.2: Ejemplo de retroalimentación formativa y la contribución de wiki evaluada

5.1.1.1.2 Método

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, basada en procedimientos de autoevaluación, evaluación entre iguales y evaluación del profesor; y una segunda parte en la que se evalúan las competencias genéricas y es donde se pone en práctica el *ciclo de contraste de hipótesis*.

5.1.1.1.3 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 profesor. 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 profesor. A continuación se describen cada una de estas fases.

Desarrollo del trabajo en el wiki

Esta fase se representa en la columna de la izquierda de la figura 5.3, y es en la que los estudiantes realizan el trabajo en las páginas del wiki. Normalmente, cada grupo de estudiantes tendrá que desarrollar su trabajos en una página del wiki. En la zona más alta de la columna se representa el comienzo del trabajo con una

5. EVALUACIÓN

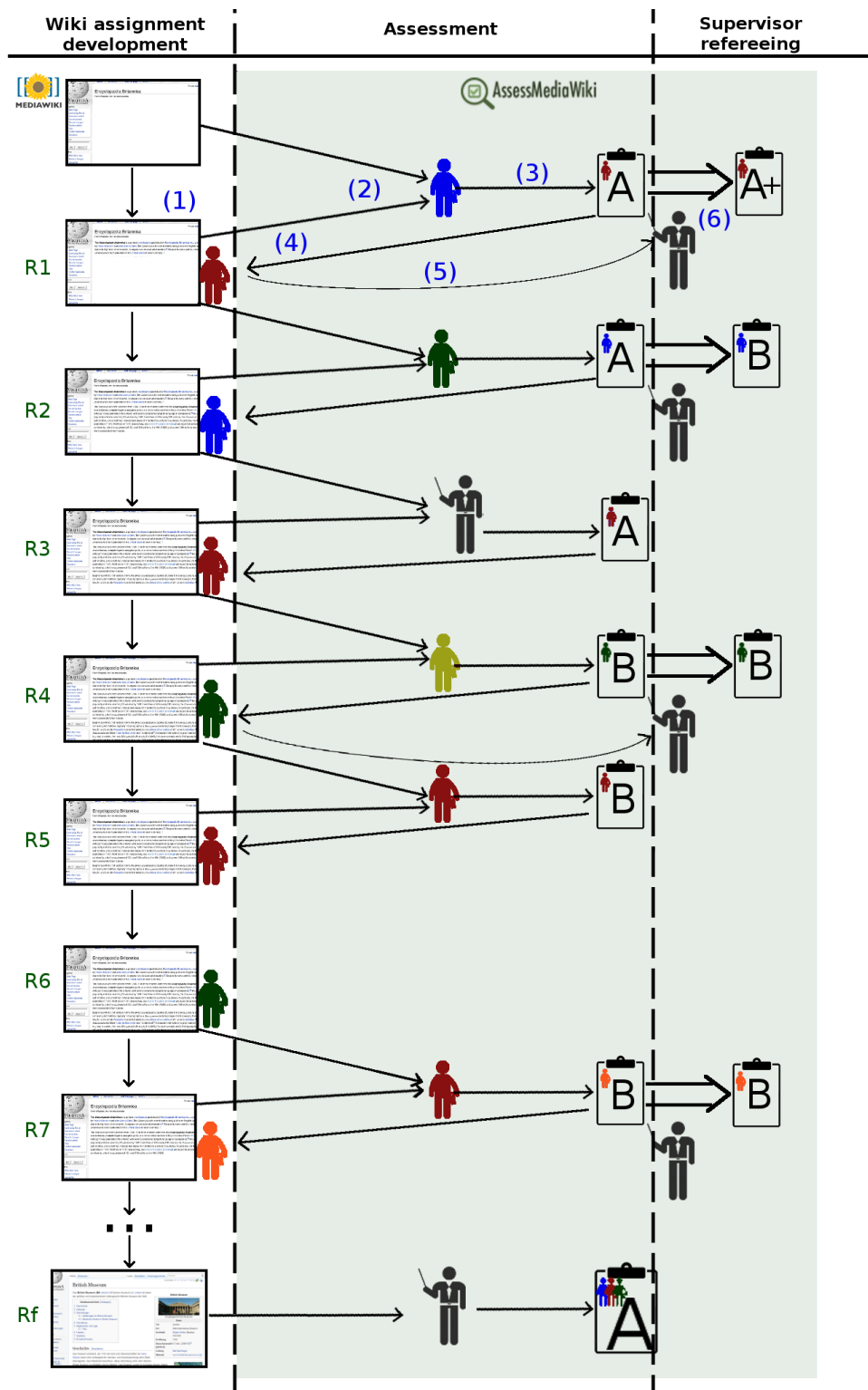


Figura 5.3: Ejemplo de flujo de trabajo para la evaluación cualitativa del wiki utilizando AMW

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 profesor 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.

Evaluación

Esta fase se muestra en la columna central y comprende las siguientes actividades:

- *Autoevaluación, evaluación entre iguales y evaluación del profesor.* 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 profesor. 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 (*1*) al estudiante azul. El estudiante azul comprueba ambas versiones para ver las diferencias entre ambas versiones (*2*) y realiza la evaluación completando la rúbrica proporcionada por el profesor (*3*). Cabe destacar también otras situaciones interesantes. En la

5. EVALUACIÓN

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 profesor el que realiza la evaluación de la contribución del estudiante rojo. Esto puede deberse a que el estudiante manualmente detecta una contribución que considera oportuno 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 profesor deberá resolver la réplica en la siguiente etapa.
- *Evaluación final del wiki.* El profesor 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 profesor 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 profesor. 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 profesor 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 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

5. EVALUACIÓN

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 profesor. De nuevo, se debería balancear el esfuerzo requerido y el detalle a exigir en las evaluaciones.

Revisión del profesor

En esta última columna se representan dos actividades que corresponden al profesor:

- Resolución de las réplicas: el profesor 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 profesor 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 profesor no acepta la réplica realizada por este último, y mantiene la calificación otorgada inicialmente por el estudiante amarillo.
- Revisión de evaluaciones no replicadas: el profesor puede revisar aleatoriamente otras evaluaciones realizadas por los estudiantes que no hayan sido replicadas. En el diagrama puede verse como el profesor revisa las evaluaciones realizadas sobre las contribuciones representadas en R2 y en R7, disminuyendo la calificación de la primera y manteniendo la segunda.

5.1.1.1.4 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 método. Esta evaluación se puede llevar a cabo durante o después de que los estudiantes

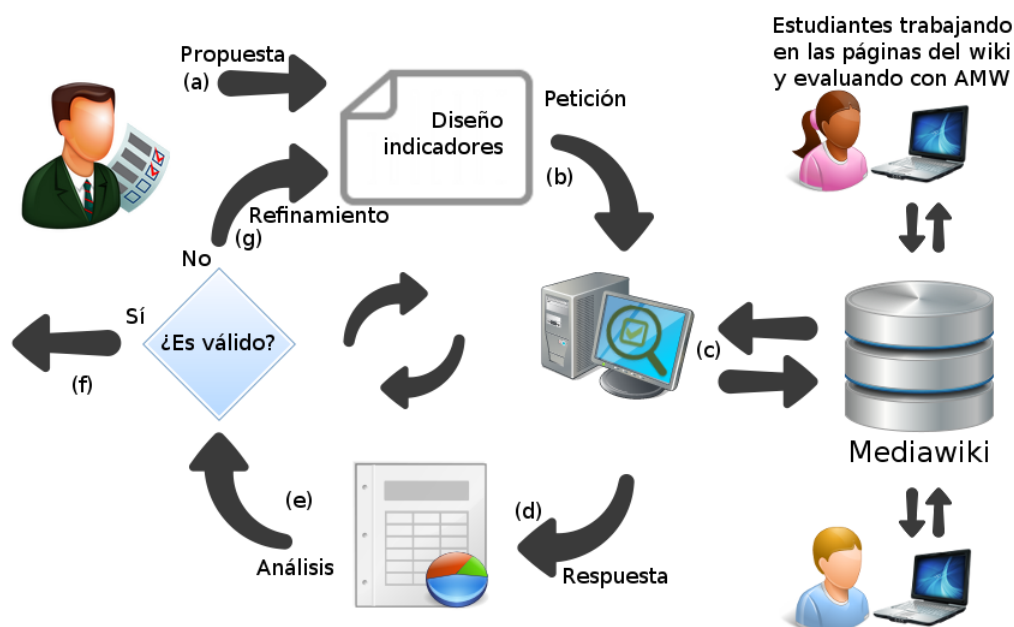


Figura 5.4: Ciclo de contraste de hipótesis para la evaluación en wikis

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 profesor que plasmará en una hoja de cálculo a partir de los datos que recibirá de las evaluaciones (a). A continuación, el profesor 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 profesor (d). El profesor 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 profesor puede rediseñar la evaluación (g) y hacer una nueva petición.

5.1.1.1.5 Indicadores

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

5. EVALUACIÓN

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 alumno 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.

5.1 Introducción

| 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 profesor | 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 profesor*. 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 profesor modificando las calificaciones, podríamos considerar que dicho alumno 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. EVALUACIÓN

5.1.1.2 VLE

En esta segunda propuesta el método se aplica a los cursos virtuales. El VLE es el núcleo de los cursos virtuales, donde los profesores ponen el material a disposición de los estudiantes y donde los estudiantes pueden fácilmente acceder a ellos; donde tanto los profesores como los estudiantes se pueden comunicar entre ellos de manera síncrona y asíncrona; y donde los estudiantes gestionan sus proyectos en equipo mientras dialogan entre ellos. Gracias a la gran cantidad de información que genera la actividad de los estudiantes y que queda registrada en el VLE, los investigadores de diferentes áreas han colaborado para extraerlos, realizar minería de datos y utilizarlos para mejorar el aprendizaje [73].

Además, en algunos de los trabajos recopilados en el estado del arte mostrado en el capítulo 2, se confirma la relación entre la interacción que llevan a cabo los estudiantes en el VLE y su rendimiento en el desempeño de varias competencias genéricas [39, 79].

Por tanto, y tal y cómo se mostrará a continuación, en esta propuesta se utilizarán indicadores obtenidos de la interacción de los estudiantes en el VLE para el diseño de evaluaciones de competencias genéricas dentro del *ciclo de contraste de hipótesis*.

5.1.1.2.1 SASQL y EvalCourse

Para diseñar evaluaciones a partir de los indicadores del VLE se crea un lenguaje específico de dominio (DSL, del inglés *domain-specific language*) [95]: *Simple Assessment-Specific Query Language* (SASQL). Es un lenguaje formal para la ejecución automática de consultas simples escritas utilizando un lenguaje específico de evaluación. SASQL tiene un sintaxis simple, orientada a la evaluación de competencias genéricas [11]. De esta forma, los profesores 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

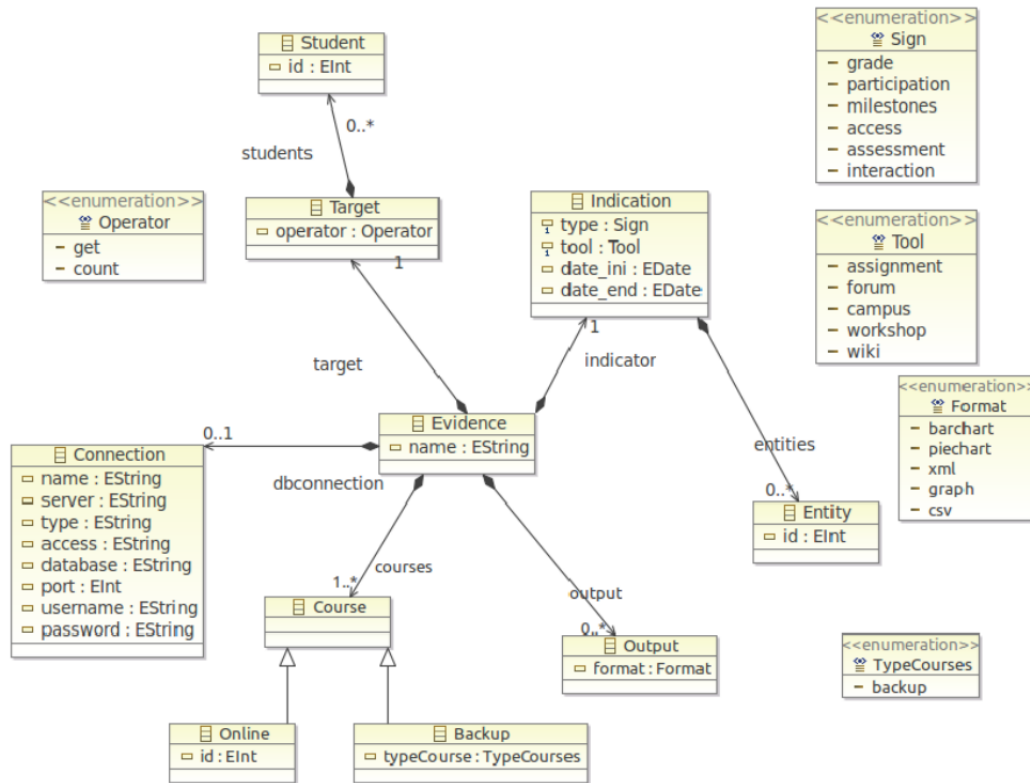


Figura 5.5: Metamodelo de SASQL

registro de actividad. EvalCourse¹ está basado en el IDE de la plataforma Eclipse, fue implementado utilizando Xtext [37] 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 EvalCourse puede verse en la figura 5.5. La entidad principal es el indicador o evidencia (*evidence*). Esta evidencia se aplicará a una herramien-

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

5. EVALUACIÓN

ta (*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*.

La sintaxis de SASQL puede verse en la consulta 5.1. 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.

```
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.
```

Consulta 5.1: Sintaxis de SASQL (las palabras reservadas se muestran resaltadas)

El objetivo es que EvalCourse pueda ser utilizado para cualquier VLE, pero esta primera versión ha sido implementada para funcionar en Moodle ¹. Moodle es

¹<https://moodle.org/>

un sistema de gestión de aprendizaje de código abierto con más de 64.000 sitios registrados ¹, y que además es el que se utiliza en la Universidad de Cádiz.

5.1.1.2.2 Método

Para evaluar las competencias genéricas el profesor deberá definir los indicadores que serán extraídos de la actividad de cada estudiante en el VLE. Ilustraremos el 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 [43].

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 profesor 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 profesor 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 profesor podrá rediseñar la evaluación (g). En cualquier caso, el profesor podrá reutilizar el diseño cuántas veces sea necesaria a lo largo del curso y monitorizar la evolución de los indicadores de cada estudiante.

Ejemplo de uso

En este ejemplo, el profesor pretende obtener indicadores del uso del foro para evaluar la competencia genérica de las habilidades interpersonales. Durante el curso,

¹<https://moodle.net/stats>

5. EVALUACIÓN

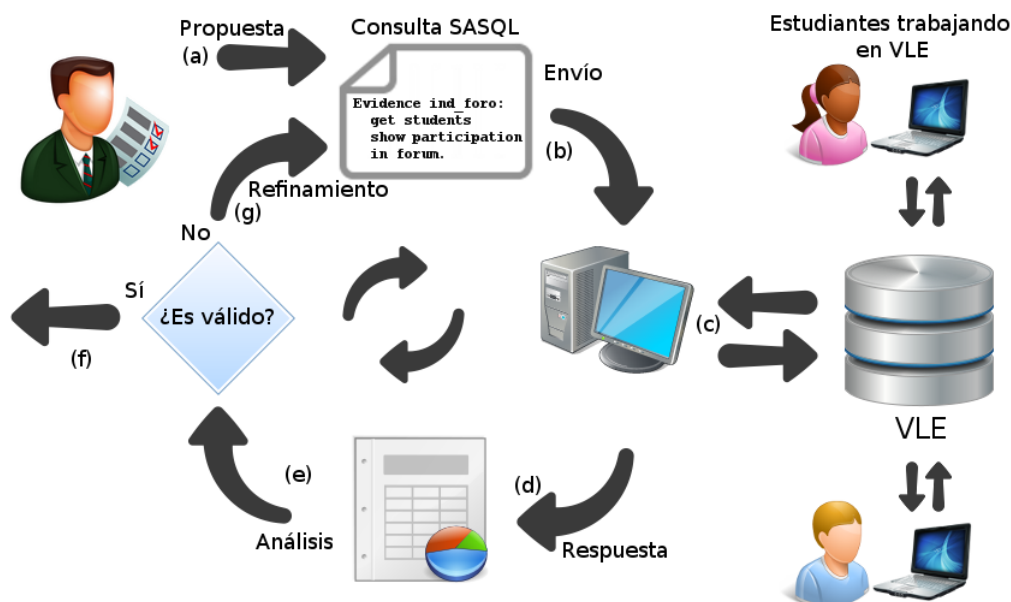


Figura 5.6: Ciclo de contraste de hipótesis utilizando EvalCourse

los estudiantes interactuarán en el foro conforme a las instrucciones proporcionadas por el profesor. Cuando lo desee, el profesor podrá utilizar EvalCourse para obtener los indicadores.

El profesor 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 5.2. EvalCourse procesa la consulta y devuelve los resultados que se pueden ver en la tabla 5.2.

```

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

Consulta 5.2: Participación en el foro en un periodo concreto de tiempo

A la vista de los resultados, y en base a la hipótesis inicial, se podría decir que todos los estudiantes menos el 3 (*student3*) habrían desempeñado correctamente la

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 |

competencia. Sin embargo, el profesor 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 5.3.

```

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

Consulta 5.3: Interacción en el foro en un periodo de tiempo

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 profesor irá redefiniendo sus hipótesis hasta dar con los indicadores que a su juicio satisfagan la evaluación de la competencia genérica.

5.1.1.2.3 Indicadores

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 profesores y no serlos para otros, así como que habrá otras muchas formas

5. EVALUACIÓN

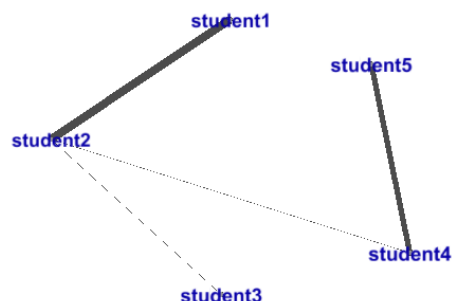


Figura 5.7: Interacción en el foro en un periodo de tiempo.

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 profesor. Por tanto, si no utilizan el foro, los estudiantes no tienen calificación en esta competencia. Por ejemplo, podría fijarse que un estudiante que tuviera tres o más intervenciones en el foro tendría una evaluación positiva en la competencia.

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

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

profesor 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 profesor 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 profesor utilizaría para cada tarea la diferencia entre la media de las notas dadas por sus compañeros y la que se asignó el propio estudiante.

Planificación y gestión del tiempo

Las tareas programadas en el VLE tienen una fecha límite de entrega. Sin embargo, el profesor puede configurar la actividad para que permita envíos retrasados. El profesor 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 profesor 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.

5.1.1.3 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 [16]. Lo mismo se puede decir de los entornos de aprendizaje como los mundos virtuales (Second Life, OpenCobalt, etc.) [52]. Esto se debe a que normalmente no se distribuyen bajo licencia libre y, por consiguiente, los profesores no pueden analizar las interacciones de los estudiantes o analizar su impacto en el aprendizaje y en los resultados de aprendizaje [27, 67].

Ante esta situación, en un trabajo previo desarrollamos nuestro propio mundo virtual, basado en OpenSim (software de código abierto) [17]. 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 .

5. EVALUACIÓN

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.

5.1.1.3.1 VWQL y EvalSim

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 [37] dentro del Eclipse Modeling Framework.

La sintaxis del lenguaje (versión beta 0.1) puede verse en la consulta 5.4. 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.

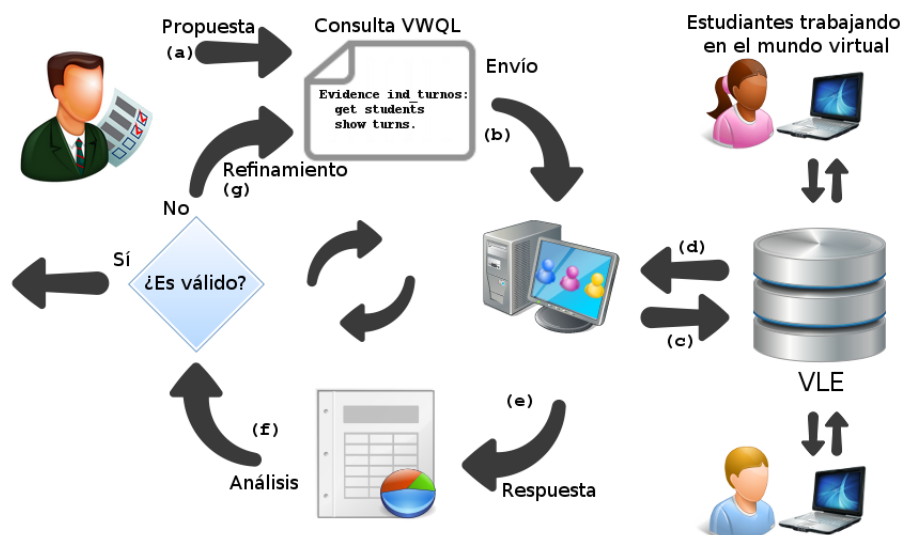


Figura 5.8: Ciclo de contraste de hipótesis con EvalSim

- *points*: número de puntos obtenidos en el mundo virtual. La manera en que los puntos se obtengan dependerá específicamente del mundo virtual jugado.

```

1 Evidence name_of_the_evidence :
2   get students [ id_of_the_student ]
3   show ( words [dict] | sentences | turns |
4         | time | points )+
  
```

Consulta 5.4: Palabras reservadas y formato de VWQL (version 0.1)

5.1.1.3.2 Método

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 profesores obtendrán indicadores del trabajo de sus estudiantes (figura 5.8).

El profesor 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

5. EVALUACIÓN

los resultados debidamente formateados al profesor (*e*). El profesor 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.

Ejemplo de uso

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 profesor, los estudiantes han de participar en un juego de roles (*role-play*) y son advertidos de que deben explayarse en sus respuestas a fin de utilizar los recursos lingüísticos aprendidos en clase.

Una vez que todos los estudiantes han participado el juego, el profesor 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 5.5.

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

Consulta 5.5: Respuestas de una sola palabra

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 profesor 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 profesor 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 profesor decide diseñar un nuevo indicador. En este caso, calculando el número de palabras por turno que escriben los estudiantes. Para ello utiliza la consulta 5.6. Los resultados en este caso puede verse en la tabla 5.4.

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 |

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

Consulta 5.6: Respuestas de una sola palabra

A la vista de los resultados, el profesor 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 profesor 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 profesor el uso que se dé a los indicadores.

5.1.1.3.3 Indicadores

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*. Aunque sólo es una propuesta y podrían utilizarse también para otras competencias.

5. EVALUACIÓN

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 |

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.

5.1.2 Publicaciones

Durante el desarrollo de este método se ha tratado de publicar en diferentes ámbitos a partir de cada una de sus implementaciones. La evaluación realizada en revistas de impacto y conferencias de relevancia son realizadas mediante procesos de *revisión por par doble-ciego*, donde tanto los revisores como los autores son anónimos [57]. De esta forma se han recibido evaluaciones, opiniones y recomendaciones, que se han ido incorporando a las diferentes implementaciones y al método.

5.1.3 Cuestionarios y entrevistas

En el transcurso de esta tesis se ha presentado en varios foros el trabajo que se va desarrollando. Los profesionales de la educación que han asistido a las presentaciones han participado en cuestionarios de evaluación de la propuesta y han vertido sus opiniones y recomendaciones sobre el método en las entrevistas realizadas. Los profesionales que han participado son profesores universitarios, profesores de primaria y secundaria y personal de MediaWiki España.

5.2 AMW aplicado a MediaWiki

El estudio de caso donde se aplica por primera vez AMW se presentó en el *SPE-DECE 2012 (Ninth multidisciplinary symposium on the design and evaluation of digital content for education)* [10] y se amplió en la versión enviada a la revista *Computers & Education* en 2015 ¹, pendiente aún de ser publicada.

¹<http://www.journals.elsevier.com/computers-and-education/>

5. EVALUACIÓN

5.2.1 Evaluación

El estudio de caso en el que se aplicó AMW tuvo lugar en la titulación de *Ingeniería Técnica en Informática de Sistemas*, de la Universidad de Cádiz, en el curso 2011-12. La experiencia fue desarrollada en la asignatura de *Administración de Sistemas Operativos* con 40 estudiantes matriculados. Dentro de las actividades de la asignatura se incluyó el desarrollo de un proyecto en el wiki, que fue evaluado mediante una aproximación mixta con SMW y AMW.

En una primera iteración del *ciclo de contraste de hipótesis* se obtuvieron indicadores cuantitativos, mientras que en la segunda iteración los indicadores que se obtuvieron fueron cualitativos. Las dimensiones que se pretendían evaluar fueron las siguientes:

- Trabajo en equipo (M_1)
- Comunicación y la aplicación del conocimiento (M_2)
- Habilidades individuales (M_3)
- Capacidad crítica (M_4)
- Entregable final (M_5)

5.2.1.1 Indicadores cuantitativos

Para la evaluación de competencias en este primer ciclo se decidió partir de indicadores cuantitativos, midiendo en bytes la cantidad de trabajo realizada por cada estudiante en las contribuciones hechas al wiki. Para la aplicación del método se usó SMW. Los indicadores que se utilizaron fueron los siguientes:

- Trabajo en equipo (M_1): Ratio de miembros del equipo que contribuyeron en el proyecto wiki.
- Comunicación y la aplicación del conocimiento (M_2): Ratio de miembros del equipo que contribuyeron al menos a un 20% del conteo final de bytes de la categoría.
- Habilidades individuales (M_3): contribución en bytes.

- Capacidad crítica (M_4): No considerado

5.2.1.2 Indicadores cualitativos

En esta segunda iteración del *ciclo de contraste de hipótesis* se utilizó AMW para obtener indicadores cualitativos. Había numerosos aspectos que se escapaban de la evaluación después de obtener los indicadores cuantitativos y que un análisis cualitativo complementaría. Los indicadores que se utilizaron en este caso para las competencias fueron los siguientes:

- Trabajo en equipo (M_1): Ratio de miembros del equipo que trabajaron en el mismo criterio técnico.
- Comunicación y la aplicación del conocimiento (M_2): Media de las notas que los miembros del equipo recibieron.
- Habilidades individuales (M_3): Media de las notas que cada estudiante recibió.
- Capacidad crítica (M_4): Número de evaluaciones que el estudiante realizó y que fueron posteriormente modificadas por el profesor.

5.2.1.3 Resumen

En la tabla 5.5 puede verse una comparativa de los indicadores cuantitativos y cualitativos. Durante el análisis de los indicadores cuantitativos se detectaron pequeños defectos que tuvieron que ser completados con indicadores cualitativos.

En la competencia de trabajo en equipo, con el enfoque cuantitativo, era muy sencillo para muchos estudiantes conseguir un indicador positivo, ya que por el simple hecho de haber escrito algo en la página era suficiente. Sin embargo, esto no justificaba en sí un trabajo en equipo, sino el haber trabajado algo, aunque fuese poco, en el mismo proyecto. Por ello se planteó el enfoque cualitativo. Al considerarse varios aspectos en un trabajo, y el estar estos aspectos representados en criterios de la rúbrica, el hecho de que dos o más estudiantes tuviesen alguna calificación en un mismo criterio podría considerarse como una evidencia de que esos

5. EVALUACIÓN

| Competencia | Indicadores cuantitativos | Indicadores cualitativos |
|--|--|--|
| Trabajo en equipo | Ratio de miembros del equipo que contribuyeron a la misma página wiki dentro del proyecto | Ratio de miembros del equipo que trabajaron en el mismo criterio técnico |
| Comunicación y aplicación del conocimiento | Ratio de miembros del equipo que contribuyeron al menos al 20% del conteo final de bytes de la categoría | Media de las notas que los miembros del equipo recibieron |
| Habilidades individuales | Contribución en bytes | Media de las notas que el estudiante recibió |
| Pensamiento crítico | <i>No considerado</i> | Número de evaluaciones que el estudiante hizo y fueron modificadas por el profesor |
| Entregable final | Manualmente | Manualmente |

Tabla 5.5: Resumen de indicadores para cada aproximación.

estudiantes han trabajado en colaboración para alcanzar el objetivo representado en dicho criterio.

Para la competencia de la comunicación y la aplicación del conocimiento se consideró la cantidad de miembros del equipo que contribuyó al menos al 20% de la versión final del wiki en términos de bytes. Este indicador evalúa el nivel de competencia de los equipos de trabajo que contribuyen al éxito del proyecto. Sin embargo, este indicador podía ser confuso. Si un estudiante mueve un trozo de texto considerable el sistema consideraría esta acción como una contribución importante cuando en realidad no lo merece. Incluso un estudiante podría engañar al sistema de pegando texto absurdo y borrándolo después, y así contribuir a unas cifras de trabajo que no serían reales pero que pasarían por alto en un análisis cuantitativo. Por eso, este indicador se complementó con un análisis cualitativo, en el que si un estudiante mueve un trozo de texto de un sitio a otro, quizás pueda obtener una buena nota en el criterio de la coherencia del texto, pero desde luego no en un criterio específico. En este enfoque cualitativo, a partir de la media de todas

las calificaciones de todas las contribuciones de los miembros del equipo se mide cómo han aplicado los estudiantes el conocimiento y cómo se han comunicado para este fin.

Para las competencias de las habilidades individuales nos encontramos ante la misma situación. Medir en bytes el trabajo de los profesores te aporta un dato cuantitativo válido, pero que necesita un análisis cualitativo que tenga en cuenta posibles situaciones que se tendrían en cuenta con el enfoque cuantitativo. Con la media de las calificaciones de las contribuciones realizadas individualmente por cada estudiante se complementa la evaluación de esta competencia.

Para el pensamiento crítico se tiene en cuenta sólo el aspecto cualitativo. Los estudiantes disponen del conocimiento para realizar la evaluación del trabajo de sus compañeros. Los estudiantes evaluados pueden reclamar una evaluación recibida si no están de acuerdo con la misma. En ese caso es el profesor el que se encarga de la revisión, y si modifica la nota, el alumno evaluador será penalizado en esta competencia.

5.2.2 Resultados

La experiencia constaba de las tres fases que fueron introducidas en el capítulo anterior:

1. Desarrollo del trabajo en el wiki Esta etapa se inició con un seminario donde los estudiantes aprendieron cómo debían trabajar colaborativamente en un wiki basado en MediaWiki. Después, los estudiantes se dividieron en doce grupos de tres miembros y dos grupos de dos miembros. Cada grupo tuvo que escribir la documentación de su proyecto en una sola página wiki. Dado que uno de los objetivos de la asignatura era desarrollar una experiencia de *evaluación auténtica*, el cometido del proyecto era la planificación y gestión del proceso de migración real de un sistema de información de la empresa. Los estudiantes fueron los responsables de la planificación y la gestión de su asignación, la coordinación de sus tareas y del trabajo en colaboración. El wiki está disponible públicamente ¹, y en seis semanas de los estudiantes

¹<http://wikis.uca.es/wikiASO/>

5. EVALUACIÓN

crearon más de 1.400 ediciones en los 14 proyectos que participaron en la asignatura. Como guía de esta experiencia, mostraremos los resultados obtenidos por los cuatro grupos de tres miembros de la tabla 5.6.

| Project (P_i) | Members (U_i) |
|-------------------|--------------------------|
| P_1 | U_7, U_8, U_9 |
| P_4 | U_{13}, U_{14}, U_{15} |
| P_7 | U_4, U_5, U_6 |
| P_{13} | U_1, U_2, U_3 |

Tabla 5.6: Grupos de ejemplo y sus miembros.

2. Evaluación de los estudiantes Esta fase comenzó con un seminario. Los estudiantes aprendieron cómo evaluar contribuciones al wiki mediante el uso de AMW. Tras ello, los estudiantes realizaron 412 evaluaciones cualitativas de contribuciones al wiki. Este proceso proporcionó a los estudiantes información crítica sobre su trabajo. Los estudiantes también pudieron replicar una evaluación si no estaban de acuerdo con la misma.

3. Revisión del profesor El profesor tuvo que resolver las réplicas. Esta información se utilizó más tarde para evaluar el pensamiento crítico.

AMW se utilizó para evaluar el desempeño de los estudiantes en varios indicadores relacionados con el trabajo en equipo. Indicadores como la coordinación para desempeñar tareas y su desempeño de manera colaborativa. Además, se fomentó el pensamiento crítico de los estudiantes motivándoles a revisar formalmente el trabajo de sus compañeros para que su nota no se bajase.

A continuación en cada subsección se explica cómo se evaluaron las competencias para cada subdimensión. Esta propuesta se basa en el programa de la asignatura. Dependiendo de la configuración de las tareas específicas del wiki y de la actividad a desarrollar, el profesor podría utilizar estos indicadores como proponemos, adaptarlos a la evaluación de otras habilidades, o incluso definir sus indicadores.

5.2.2.1 Indicador del trabajo en equipo (M_1)

Aproximación cuantitativa

Mediante el uso de SMW se pudo comprobar como todos los miembros de cada equipo, en menor o mayor medida habían participado en la páginas wikis de sus proyectos. Este indicador se descartó, ya que una vez analizado lo único que podíamos saber era si los estudiantes habían contribuido al proyecto, pero el hecho de haber contribuido al mismo proyecto no es necesariamente una evidencia de haber trabajado en equipo. Por tanto, este indicador se redefinió mediante la aproximación cualitativa.

Aproximación cualitativa

Esta calificación mide la habilidad de un equipo para trabajar colaborativamente y era compartida por todos los miembros del equipo (U_i) de cada proyecto (P_i). Esta dimensión recogía los criterios en los que fue evaluado cada U_i en la página de su proyecto. Consideramos que un U_i ha contribuido en un criterio de un proyecto si ha sido evaluado en ese criterio en alguna de sus contribuciones.

De esta forma, dos o más estudiantes colaboraron si fueron evaluados en el mismo criterio. Si dos miembros del equipo fueron evaluados en criterios complementarios, ello no implicaba colaboración. Bajo esta premisa sólo se consideraron los criterios técnicos, que están relacionados con las partes de la tarea en la que los estudiantes en realidad colaboraron. La colaboración en criterios transversales como la escritura o la coherencia no se consideraron (D_{11} - D_{17}).

En la tabla 5.7 se muestran los criterios evaluados para cada miembro del proyecto P_{13} . Mientras que U_1 fue evaluado en cinco criterios, U_2 y U_3 sólo fueron evaluados de uno. Los criterios *servidor físico* (D_4), *red* (D_6), *PCs* (D_7), *Gantt* (D_9) y *presupuesto* (D_{10}) sólo fueron trabajados por un miembro del equipo (U_1). El único criterio trabajado por más de un miembro del equipo fue D_8 , que fue trabajado por U_1 y U_2 . Así que, bajo nuestra propuesta, U_1 y U_2 trabajaron colaborativamente, pero U_3 no.

Finalmente, la calificación del grupo en esta dimensión se calculaba con la fórmula 5.1.

5. EVALUACIÓN

| Estudiante | D_1 | D_2 | D_3 | D_4 | D_5 | D_6 | D_7 | D_8 | D_9 | D_{10} | Colaboración |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------------------|
| U_1 | | | | 1 | | | 1 | 1 | 1 | 1 | sí (en D_8 con U_2) |
| U_2 | | | | | | | | 1 | | | sí (en D_8 con U_1) |
| U_3 | | | | | | 1 | | | | | no |

Tabla 5.7: Grading of collaborative work (M_1) in the P_{13} project.

$$M_1 = CTM/TM \quad (5.1)$$

En dicha fórmula participan las siguientes variables:

- CTM (*collaborative team members*): número de miembros del equipo que trabajaron colaborativamente.
- TM (*team members*): número de miembros que componen el equipo.

La calificación M_1 para el proyecto P_{13} fue de 6,66 sobre 10, ya que dos de los tres miembros trabajaron colaborativamente.

5.2.2.2 Indicador de las habilidades de comunicación y la aplicación del conocimiento (M_2)

Aproximación cuantitativa

En este enfoque se considera la cantidad de miembros de cada equipo que habían contribuido al menos a un 20% del trabajo realizado en la página del wiki medido en bytes (para equipos de tres estudiantes). En la figura 5.9 puede verse una gráfica de la distribución del trabajo, mostrándose el ratio total de bytes contribuidos por cada estudiante del proyecto P_{13} . U_1 (verde) contribuyó con un 38,2%, U_2 (rojo) con un 27,2% y U_3 (azul) con un 34,7%. Al estar más o menos balanceada consideramos que ellos trabajaron colaborativamente.

Aproximación cualitativa

Esta calificación medía la habilidad para aplicar el conocimiento en situaciones prácticas y la habilidad para comunicarse con otros miembros del equipo. La calificación de este indicador también era compartida por todos los miembros del

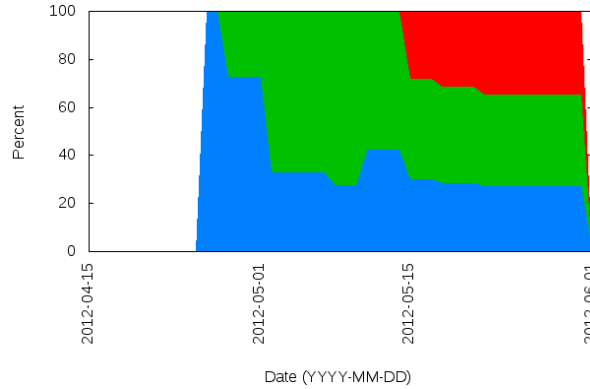


Figura 5.9: Distribución del trabajo de los miembros del proyecto P_{13} .

equipo (U_i). Se calcula como la media de las notas que cada miembro del equipo recibió. Este indicador evaluaba la eficacia de los equipo de trabajo que contribuyen al éxito del proyecto. Notas medias en las contribuciones pueden significar pobres contribuciones al wiki o que algunas contribuciones obtuvieron buenas calificaciones y otras malas, evidenciando una comunicación deficiente entre los miembros del equipo o un escaso grado de compromiso con el equipo.

La media de todas las calificaciones dadas a los criterios de cada contribución a los miembros de cada equipo en el proyecto se calcula mediante la fórmula 5.2.

$$M_2 = SRA_G / NRA_G \quad (5.2)$$

En dicha fórmula participan las siguientes variables:

- SRA_G (*sum of received assessments*): suma de las calificaciones recibidas para todas las contribuciones del proyecto.
- NRA_G (*number of received assessments*): número de evaluaciones recibidas.

5.2.2.3 Indicador de habilidades individuales (M_3)

Aproximación cuantitativa

bla bla bla

5. EVALUACIÓN

| Estudiante | D_1 | D_2 | D_3 | D_4 | D_5 | D_6 | D_7 | D_8 | D_9 | D_{10} | D_{11} | SRA_G | NRA_G |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|-------------|-----------|
| U_4 | 15 | 13 | 7 | 26 | 22 | 34 | 30 | 6 | 25 | 37 | 89 | 304 | 40 |
| U_5 | | | | 26 | 8 | | | | 18 | | 17 | 69 | 8 |
| U_6 | | 8 | | 25 | 9 | | 7 | 8 | | 8 | 133 | 198 | 26 |
| Suma | | | | | | | | | | | | 571 | 74 |
| Media | | | | | | | | | | | | 7.71 | |

Tabla 5.8: (M_2) para el proyecto P_7 .

| Estudiante | D_1 | D_2 | D_3 | D_4 | D_5 | D_6 | D_7 | D_8 | D_9 | D_{10} | D_{11} | SRA_G | NRA_G |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|-------------|-----------|
| U_{13} | | | | 24 | | 15 | 7 | 7 | 8 | 47 | 98 | 206 | 27 |
| U_{14} | | | | 26 | 8 | | | | 27 | | 10 | 37 | 4 |
| U_{15} | 22 | 46 | 21 | 40 | 17 | 5 | 94 | 5 | 6 | 20 | 123 | 399 | 64 |
| Suma | | | | | | | | | | | | 642 | 95 |
| Media | | | | | | | | | | | | 6.75 | |

Tabla 5.9: M_2 para el proyecto P_4 .

Aproximación cualitativa

Esta dimensión evalúa la capacidad para producir y mantener la calidad del proyecto. Es una habilidad individual, y se calcula por separado para cada miembro del equipo (U_i). El grado en esta dimensión es el promedio de las calificaciones recibidas para cada alumno, expresado en la fórmula 5.3.

$$M_3 = SRA_S / NRA_S \quad (5.3)$$

En dicha fórmula participan las siguientes variables:

- SRA_S (*sum of received assessments by the student*): suma de las calificaciones recibidas para las contribuciones del estudiante.
- NRA_S (*number of received assessments by the student*): número de evaluaciones recibidas por el estudiante.

El promedio de las calificaciones asignadas a las contribuciones de cada estudiante del proyecto P_4 (U_{13} , U_{14} , U_{15}) y del proyecto P_{13} (U_1 , U_2 , U_3) puede verse

5.2 AMW aplicado a MediaWiki

en la tabla 5.10. Bajo nuestro enfoque, los miembros de P_{13} tuvieron calificaciones similares. Pero en P_4 puede verse un contraste, ya que U_{14} hizo buenas aportaciones, mientras que U_{15} no contribuyó a la calidad del proyecto.

| P_i | U_i | D_1 | D_2 | D_3 | D_4 | D_5 | D_6 | D_7 | D_8 | D_9 | D_{10} | D_{11} | SRA_G | NRA_G | M |
|----------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------|-----------|-----|
| P_{13} | U_1 | | | | 7 | | | 15 | 13 | 16 | 40 | 24 | 115 | 15 | 7.6 |
| | U_2 | | | | | | | | 8 | | | | 8 | 1 | 8.0 |
| | U_3 | | | | | | 14 | | | | | 15 | 29 | 4 | 7.2 |
| P_4 | U_{13} | | | | 24 | | 15 | 7 | 7 | 8 | 47 | 98 | 206 | 27 | 7.6 |
| | U_{14} | | | | 26 | 8 | | | | 27 | | 10 | 37 | 4 | 9.2 |
| | U_{15} | 22 | 46 | 21 | 40 | 17 | 5 | 94 | 5 | 6 | 20 | 123 | 399 | 64 | 6.2 |

Tabla 5.10: (M_3) para los miembros de P_4 y P_{13} .

5.2.2.4 Indicador del pensamiento crítico (M_4)

Aproximación cuantitativa

bla bla bla

Aproximación cualitativa

Como se ha comentado anteriormente, AMW permitió a los estudiantes replicar si estaban en desacuerdo con una evaluación recibida. Esas réplicas se reportaron al supervisor para resolverlos. En caso de que el profesor aceptase una réplica, la calificación se modificaba. Pero este proceso tiene otra parte: los estudiantes recibieron instrucciones claras sobre la tarea de evaluación, siendo un proceso formativo que tenía reflejo en su calificación. Entonces, podíamos utilizar esta información como un indicador de la habilidad para *interpretar, analizar y evaluar el trabajo de sus compañeros*. En particular, hemos utilizado dos indicadores: el número de evaluaciones realizadas por cada estudiante y la cantidad de ellos que el profesor modificó (podían tener su origen en réplicas del estudiante evaluado o en revisiones al azar realizadas por el profesor). No fue nuestro caso, pero si había diferentes evaluaciones en una misma contribución, la diferencia entre ellos podría ser también un indicador de competencia.

5. EVALUACIÓN

La fórmula que se utilizó fue 5.4. Se considera que todos los estudiantes comenzaron con 10 puntos en esta dimensión, y perdieron 2,5 de ellos por cada evaluación que hicieron y que fue modificado por el profesor, hasta un mínimo de 0 puntos. De hecho, cada estudiante hizo 10 evaluaciones, por lo que consideramos que deberían haber hecho no más del 20% de las evaluaciones erróneas para aprobar esta parte de la nota. En nuestro caso de estudio, tuvimos 412 evaluación, y 27 de ellas (6%) se replicaron (el profesor sólo revisó las evaluaciones replicadas). Sólo 10 de ellas (37%) fueron aceptadas. En cuanto a los estudiantes, sólo uno tuvo 3 réplicas a sus evaluaciones aceptadas. El resto de ellos tenía 2, 1 o ninguno. En la tabla 5.11 pueden verse las evaluaciones realizadas por los miembros de P_{13} que fueron respondidas. U_1 y U_2 recibieron una réplica que fue aceptada (AR) cada uno, mientras que U_3 no tuvo ninguna.

$$M_4 = 10 - 2,5 * AR \quad (5.4)$$

| Students | Received replies | Accepted replies | M_4 |
|----------|------------------|------------------|-------|
| U_1 | 2 | 1 | 7.50 |
| U_2 | 1 | 1 | 7.50 |
| U_3 | 0 | 0 | 10.00 |

Tabla 5.11: (M_4) para los miembros del proyecto P_{13} .

5.2.2.5 Entregable final (M_5)

Para la calificación fina se tomó el trabajo final del wiki, donde se valoraron todos los requisitos que se esperaban del trabajo. En nuestro enfoque, cada P_i tuvo una calificación entre 0-10. Esta fue la calificación dada por el profesor una vez se ha llegado a la fecha límite. La calificación fue la misma para todos los miembros del equipo (U_i), ya que todos ellos (como equipo) son responsables del resultado final del proyecto. Los proyectos fueron evaluados usando la rúbrica indica en el cuadro 5.12. Hay dos categorías de criterios en las rúbricas:

- Criterios específicos: atributos desde D_1 a D_{10} , que se refieren a aspectos específicos del trabajo.

5.2 AMW aplicado a MediaWiki

- Criterios transversales: atributos desde D_{11} a D_{17} , que se refieren a aspectos transversales del trabajo: coherencia, correcto uso del formato wiki, ... etc.

| Item | Criterios | Peso | Nota (P_{13}) |
|-------------|----------------------------------|---------------------|-------------------|
| D_1 | Justificación | 2.00 | 2.00 |
| D_2 | Centro de datos anterior | 0.25 | 0.25 |
| D_3 | Centro de datos nuevo | 0.25 | 0.25 |
| D_4 | Servidor físico | 0.25 | 0.25 |
| D_5 | Servidor virtual | 0.25 | 0.25 |
| D_6 | Red | 0.50 | 0.50 |
| D_7 | PCs | 0.40 | 0.40 |
| D_8 | Formación | 0.25 | 0.25 |
| D_9 | Diagrama de Gantt | 0.40 | 0.20 |
| D_{10} | Presupuesto | 0.40 | 0.40 |
| D_{11} | Escritura | 0.30 | 0.30 |
| D_{12} | Coherencia | 3.00 | 2.40 |
| D_{13} | Referencias | 0.25 | 0.00 |
| D_{14} | Formato wiki | 0.50 | 0.50 |
| D_{15} | Aplicación de conceptos teóricos | 1.00 | 0.50 |
| D_{16} | Información extra | 1.00 | 0.00 |
| D_{17} | Plagio | -10.00 | -0.00 |
| Suma | | (Máx.) 10.00 | 8.45 |

Tabla 5.12: Calificaciones dadas por el profesor al proyecto P_{13} .

A modo de ejemplo, la cuarta columna de la tabla muestra las calificaciones que el supervisor le dio a los criterios del proyecto P_{13} . Este proyecto fue calificado con un 8,45 puntos sobre 10. Esta evaluación cualitativa manual realizada por el supervisor es escalable, ya que cada grupo solo tiene una versión definitiva de su página wiki.

5. EVALUACIÓN

5.2.3 Discusión

Junto con el estudio de caso, varias comparaciones entre el enfoque cuantitativo previo y la cualitativa implementada en esta experiencia han sido desplegados. Un resumen sobre las ventajas del enfoque cualitativo se muestra a continuación:

- El enfoque cualitativo permite un análisis más fino de grano de trabajo colaborativo en comparación con los que en el experimento cuantitativo anterior.
- Además, este análisis también puede detectar a los estudiantes que colaboran con un objetivo común en las diferentes páginas del wiki de un mismo proyecto.
- Este enfoque permite la detección de las contribuciones que copiar, borrar y pegar de grandes fragmentos de texto sin tener que mejorar la página wiki, lo que podría tener una buena medición cuantitativa inmerecida.
- Las contribuciones a otros proyectos se incluyen en la calificación individual de los estudiantes, mientras que estas contribuciones fueron de grano grueso en cuenta en el enfoque cuantitativo anterior.
- La capacidad de los estudiantes para ser crítico es entrenado y evaluado por el proceso de respuesta.

Mientras que las principales desventajas son las siguientes:

- La función de selección debe ser cuidadosamente elegido, por lo que las evaluaciones se llevan a cabo en las contribuciones significativas.
- Revisión de todas las evaluaciones de los compañeros y auto realizados por los estudiantes sigue siendo una tarea difícilmente escalable para el supervisor. Por lo tanto, algunas evaluaciones pobres pueden no detectados por el supervisor si no se informan, ni eligieron al azar para ser revisado.

En líneas generales, los estudiantes tuvieron un buen desempeño en la tarea. En la tabla 5.13 puede verse como solo 3 estudiantes no aprobaron el proyecto (6,98 %). Sin embargo, 37 estudiantes (86,04 %) tuvieron una buena calificación,

5.3 EvalCourse aplicado a entornos de aprendizaje virtual

más de 7 sobre 10. Los estudiantes llevaron a cabo tareas de análisis crítico mediante la evaluación del trabajo de sus compañeros. Así, obtuvieron retroalimentación y evidencias para cada evaluación recibida.

| Notas | Estudiantes | Ratio |
|--------|-------------|---------|
| [9-10] | 16 | 37.21 % |
| [8-9) | 10 | 23.25 % |
| [7-8) | 11 | 25.58 % |
| [6-7) | 3 | 6.98 % |
| [5-6) | 0 | 0 % |
| [0-5) | 3 | 6.98 % |

Tabla 5.13: Resumen de calificaciones finales de los estudiantes

5.3 EvalCourse aplicado a entornos de aprendizaje virtual

Los estudios de caso que se presentan en este capítulo fueron publicados en la revista *International Journal of Engineering Education, Special issue on Innovative Methods of Teaching Engineering* [9], después de haber sido invitados a extender un primer trabajo presentado en el *4th International Workshop on Software Engineering for E-learning (ISELEAR'13)* [12].

5.3.1 Evaluación

Los estudios de caso en los que se aplicó EvalCourse tuvieron lugar en dos asignaturas de la titulación de *Ingeniería Informática*, de la Universidad de Cádiz, en el curso 2012/13. La evaluación del curso se realizó de forma manual, y después se aplicó EvalCourse para evaluar a los estudiantes en las competencias de liderazgo, habilidades interpersonales y el pensamiento crítico.

5. EVALUACIÓN

5.3.1.1 Extracción de indicadores del foro

El primero de los casos de estudio tuvo lugar en la asignatura de *Procesadores de Lenguajes II* y que tenía 36 estudiantes matriculados. Era una asignatura obligatoria, que tenía lugar en el primer semestre del quinto (y último) curso. 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. Este foro se utilizó para evaluar dos competencias: habilidades interpersonales y liderazgo.

En primer lugar, como indicador de las habilidades interpersonales de cada estudiante se calculó el total de intervenciones en el foro. El coordinador del curso animó a sus estudiantes a participar en el foro del equipo, ya que debía ser su herramienta de comunicación interna. Un estudiante que tenía tres o más intervenciones en el foro tendría una buena calificación. En segundo lugar, como indicador de liderazgo se consideró los debates que los estudiantes iniciaron en el foro. Un estudiante que inició dos o más debates tuvo una calificación positiva en la competencia de liderazgo.

Análisis

Sólo 18 de los 36 estudiantes del curso participaron en los foros. No se puede determinar que los estudiantes que no participaron en el foro no posean las competencias, lo único que podemos afirmar es que no demostraron su desempeño en estas competencias a través de esta actividad. Los resultados de haber aplicado EvalCourse pueden verse en el listado de la tabla 5.14. Pocos estudiantes obtuvieron una calificación positiva en estas competencias, habiendo varios estudiantes que destacan sobre el resto. Estos datos se exportaron a una hoja de cálculo y se generó la figura 5.10. En resumen, consideramos que los estudiantes, a pesar de que sabían que debían utilizar el foro para la comunicación entre los miembros del equipo, no lo hicieron en su mayoría.

A pesar de esto, de manera informal sí se observó que los estudiantes con calificaciones positivas en ambas competencias realmente tuvieron un buen desempeño de dichas competencias en la asignatura. Por tanto, pudimos decir que los indicadores elegidos parecían ser verdaderas evidencias del nivel de competencia de los

5.3 EvalCourse aplicado a entornos de aprendizaje virtual

Tabla 5.14: Listado de intervenciones de los estudiante en el foro.

| id | username | Debate- starter | Debate- participation | Total |
|----|----------|--------------------|--------------------------|-------|
| 43 | S2 | 0 | 1 | 1 |
| 44 | S3 | 3 | 4 | 7 |
| 45 | S4 | 2 | 2 | 4 |
| 46 | S5 | 0 | 1 | 1 |
| 48 | S7 | 4 | 3 | 7 |
| 50 | S9 | 6 | 8 | 14 |
| 51 | S10 | 1 | 1 | 2 |
| 53 | S12 | 0 | 1 | 1 |
| 55 | S14 | 1 | 1 | 2 |
| 59 | S18 | 1 | 2 | 3 |
| 60 | S19 | 2 | 0 | 2 |
| 61 | S20 | 0 | 2 | 2 |
| 62 | S21 | 0 | 1 | 1 |
| 67 | S26 | 0 | 1 | 1 |
| 70 | S29 | 0 | 1 | 1 |
| 71 | S30 | 2 | 1 | 3 |
| 75 | S34 | 1 | 5 | 6 |
| 78 | S37 | 2 | 4 | 6 |

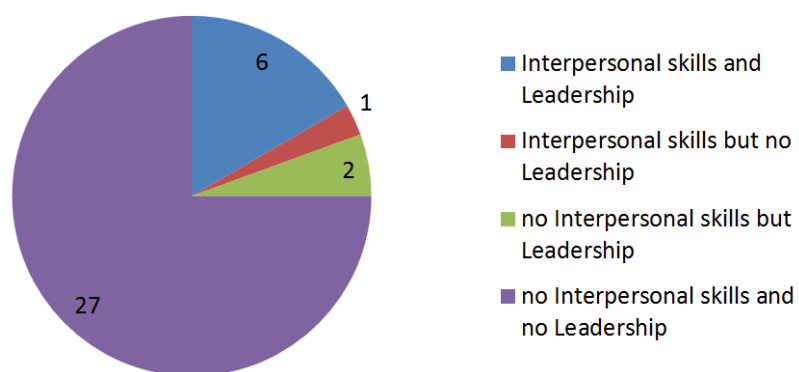


Figura 5.10: Resumen del desempeño de los estudiantes en el foro en las dos competencias.

5. EVALUACIÓN

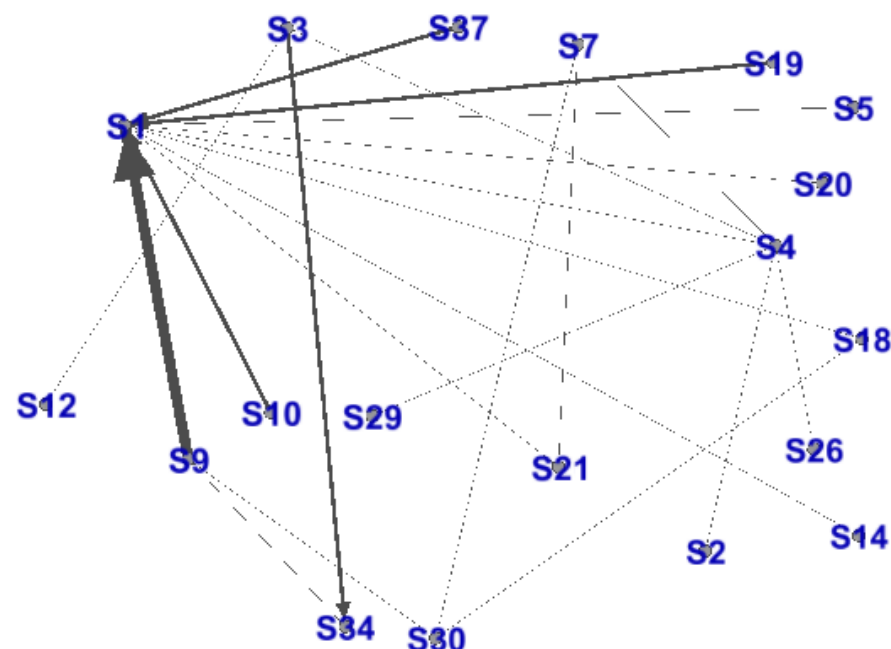


Figura 5.11: Graph of students' interaction in forum case study.

estudiantes. Probablemente el hecho de que el valor dado a la participación en el foro en la calificación global de la asignatura fuera sólo de un 2,5% no animó a los estudiantes a no utilizar en exclusiva sus habituales métodos de comunicación (Whatsapp, correo electrónico, reuniones en el pasillo, etc.).

También se utilizó EvalCourse para analizar la interacción entre los estudiantes. La idea era localizar interacciones que quizás estuviese más interesadas en mejorar sus resultados que en la mera comunicación. Se detectaron evidencias de interacciones fraudulentas entre dos estudiantes cuando se podía ver que ellos sólo estaban interesados en hablar entre ellos dos y no habían interactuado con ningún otro compañero. Con el código SASQL mostrado en la consulta 5.7 se obtuvo la interacción que hubo en el foro. Los resultados se pueden ver en la figura 5.11.

```
1 Evidence forum_interaction :  
2     get students  
3     show interaction  
4     in forum .
```

Consulta 5.7: Código SASQL para extraer datos sobre la interacción en el foro

5.3 EvalCourse aplicado a entornos de aprendizaje virtual

Concluimos que gran parte de la interacción del foro se debía a mensajes de respuesta al profesor (S1). Esto hecho es un indicio de que en muchos casos la interacción de los estudiantes con el foro es más debida a cuestiones que los estudiantes realizan al profesor para que les aclaren algunas de las instrucciones proporcionadas en el propio foro que al uso del foro como herramienta del grupo.

5.3.1.2 Extracción de indicadores del taller

El segundo estudio de caso que se analizó se desarrolló en la asignatura de Programación Funcional, en la que había 19 estudiantes matriculados. Esta es una asignatura optativa que tiene lugar también en el segundo semestre del quinto (y último) curso. 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 profesor, puede ser evaluado por otro u otros estudiantes o auto-evaluado. Los estudiantes debían entregar una tarea en un taller habilitado para cada tema antes de la fecha límite fijada en la actividad. Tras la fecha de entrega, el profesor proporcionaba la solución a la tarea, de manera que cada tarea fuera evaluada por dos compañeros de manera aleatoria y por el propio estudiante. A final de curso, se calculó la media de las calificaciones de todos los talleres. Esta calificación era el 30% de la nota de la asignatura, siendo obligatoria una buena calificación para aprobar la asignatura. El profesor de la asignatura tenía que evaluar el pensamiento crítico de sus estudiantes. Para ello, utilizó como indicador para cada tarea la diferencia entre la media de las calificaciones dadas por cada compañero con respecto a la calificación que se asignó el propio estudiante. Cada estudiante evaluaba su propio trabajo antes de saber la calificación que le habían dado sus compañeros.

El grado de validez del indicador proporcionado por EvalCourse dependerá del profesor, que si lo considera oportuno, podrá contrastar o complementar el indicador con otras actividades de aprendizaje. Al igual que los estudiantes deben haber sido instruidos para realizar el trabajo que se les pedía, deben ser capaces de evaluar el trabajo de sus compañeros, argumentando sus criterios, sus razones y sus evidencias.

5. EVALUACIÓN

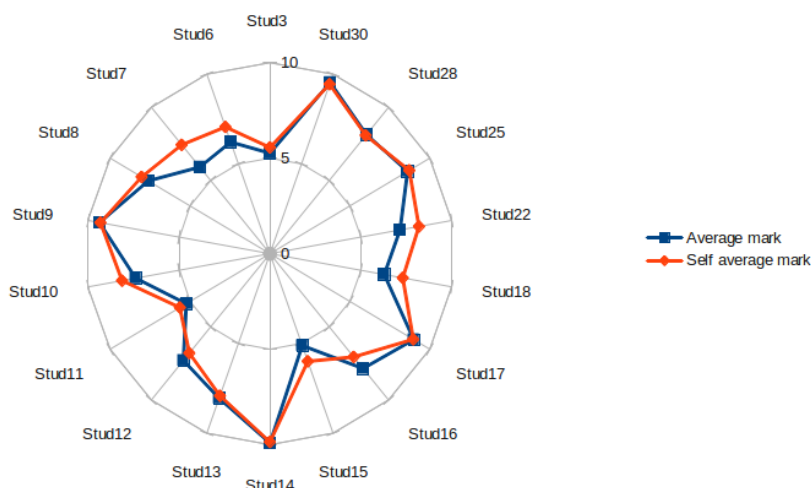


Figura 5.12: Grafo que muestra las diferentes calificaciones que se autoasignaron los estudiantes y la que les dieron sus compañeros.

Análisis

A partir del código mostrado en la consulta 5.8, se obtiene el listado 5.15 con los indicadores que se utilizaron para evaluar a los estudiantes en la competencia del pensamiento crítico (en un rango de 0 a 10). En la figura 5.12 se muestra un grafo con la diferencia entre sus auto-calificaciones y las otorgadas por sus compañeros. Este grafo se generó a partir de la importación de los indicadores a una hoja de cálculo. En él, se detecta fácilmente a dos estudiantes que se auto-asignaron una calificación inferior a la que le dieron sus compañeros.

```
1 Evidence workshop_critical :  
2     get students  
3     show evaluations  
4     in workshop .
```

Consulta 5.8: Extracción evaluaciones realizadas en los talleres

Se puede decir que no hubo muchas diferencias en las calificaciones. Esto es un buen indicador, ya que la mayor parte de los estudiantes que tuvieron una diferencia grande en las primeras calificaciones, corrigieron esta desviación en las siguientes evaluaciones. Aunque también hay estudiantes que hicieron buenas auto-evaluaciones desde el principio, como se puede comprobar viendo la última colum-

5.3 EvalCourse aplicado a entornos de aprendizaje virtual

Tabla 5.15: Desempeño de los estudiantes en la competencia de la autocrítica.

| username | Peer-grade | Self-grade | Mean-diff |
|----------|------------|------------|-----------|
| Stud3 | 5.26 | 5.56 | 0.3 |
| Stud6 | 6.22 | 7.06 | 0.94 |
| Stud7 | 5.92 | 7.44 | 1.52 |
| Stud8 | 7.62 | 8.03 | 0.60 |
| Stud9 | 9.38 | 9.33 | 0.57 |
| Stud10 | 7.39 | 8.14 | 0.75 |
| Stud11 | 5.24 | 5.64 | 0.62 |
| Stud12 | 7.30 | 6.80 | 0.83 |
| Stud13 | 8.06 | 7.92 | 0.55 |
| Stud14 | 9.92 | 9.86 | 0.22 |
| Stud15 | 5.12 | 6.00 | 1.43 |
| Stud16 | 7.85 | 7.06 | 0.79 |
| Stud17 | 9.00 | 8.97 | 0.75 |
| Stud18 | 6.29 | 7.31 | 1.01 |
| Stud22 | 7.16 | 8.20 | 1.03 |
| Stud25 | 8.61 | 8.72 | 0.22 |
| Stud28 | 8.14 | 8.08 | 0.47 |
| Stud30 | 9.54 | 9.44 | 0.23 |

5. EVALUACIÓN

na de la tabla 5.15. En ella se representa la diferencia media de las calificaciones que recibieron y que se autoasignaron.

En este punto el profesor tuvo que decidir sobre la validez del indicador. Y en caso afirmativo, cuál sería el límite entre una valoración positiva y una negativa. Gracias a que el grupo era pequeño, el profesor podía comparar el comportamiento real observado durante el curso con los resultados del indicador, confirmando que había mucha similitud entre ambas calificaicones. De hecho, el límite positivo se estableció con una media de 0,5 puntos o menos. Mientras que una valoración negativa se dejó para aquellos cuya diferencia era superior a 1. Con estos resultados se obtuvo la siguiente distribución (provocando una distribución de campana).

- Si la diferencia media está entre 0 y 0,5, esto era un indicador positivo del desempeño de la competencia de la autocrítica. (5 estudiantes).
- Si la diferencia media está entre 0,5 y 1, los estudiantes mostraron un nivel medio de la competencia (9 estudiantes).
- If the mean difference of the averages is 1 or higher, esto era un indicador negativo del desempeño de la competencia de la autocrítica (4 estudiantes).

Por supuesto, se necesitan más estudios para determinar la validez de los resultados de EvalCourse como indicadores de la competencia. Sin embargo, en nuestro caso, pudimos decir que la informaicón fue de utilidad para justificar la evaluación de los estudiantes.

5.3.2 Resultados

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.

5.4 EvalSim aplicado a los mundos virtuales

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.

5.5 Evaluación

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.

5.6 Conclusiones

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,

5. EVALUACIÓN

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.

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

[?]

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

6. CONCLUSIONES Y TRABAJO FUTURO

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.

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, 71
- [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. 71
- [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. 71, 95, 117
- [10] 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. 71, 103
- [11] 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. 71, 90
- [12] A. Balderas, I. Ruiz-Rube, M. Palomo-Duarte, and J. M. Dodero. A generative computer language to customize online learning assessments. In *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, pages 141–147. ACM, 2013. 71, 117
- [13] 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

- [14] B. Bell and B. Cowie. The characteristics of formative assessment in science education. *Science education*, 85(5):536–553, 2001. 29
- [15] 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
- [16] A. Berns, A. Gonzalez-Pardo, and D. Camacho. Game-like language learning in 3-D virtual environments. *Computers & Education*, 60(1):210–220, 2013. 97
- [17] 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. 97
- [18] A. Carreras Marín, Y. Blasco, M. Badia-Miró, M. Bosch Príncipe, I. Morillo, G. Cairó i Céspedes, and D. Casares Vidal. The promotion and assessment of generic skills from interdisciplinary teaching teams. *EDULEARN13 Proceedings*, pages 201–207, 2013. 39, 40, 41, 51, 54, 62
- [19] 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
- [20] 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
- [21] 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

BIBLIOGRAFÍA

- [22] 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
- [23] C.-h. Chen. The implementation and evaluation of a mobile self-and peer-assessment system. *Computers & Education*, 55(1):229–236, 2010. 23
- [24] 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
- [25] B. Communiqué. Making the most of our potential: Consolidating the european higher education area. In *European Higher Education Area Ministerial Conference*, volume 16, page 2012, 2012. 3
- [26] 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
- [27] 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. 97
- [28] D. D. Curtis. The assessment of generic skills. *Generic skills in vocational education and training: Research findings*, pages 136–56, 2004. 1
- [29] B. De La Harpe, A. Radloff, and J. Wyber. Quality and generic (professional) skills. *Quality in Higher Education*, 6(3):231–243, 2000. 1
- [30] 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

- [31] H. Dixon. What is teamwork? *The Canadian Journal of Medical Radiation Technology/CAMRT*, 23(2):79–80, 1992. 28
- [32] 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
- [33] 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
- [34] I. Elgort, A. G. Smith, and J. Toland. Is wiki an effective platform for group course work? *Australasian Journal of Educational Technology*, 24(2):195–210, 2008. 78
- [35] 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
- [36] 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
- [37] M. Eysholdt and H. Behrens. Xtext: implement your language faster than the quick and dirty way. In *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, pages 307–309. ACM, 2010. 91, 98
- [38] 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

BIBLIOGRAFÍA

- [39] 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, 90
- [40] A. Fisher. *Critical thinking: An introduction*. Cambridge University Press, 2011. 28
- [41] 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
- [42] M. Fowler. *Domain-specific languages*. Pearson Education, 2010. 75
- [43] 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. 93
- [44] S. M. Gass. *Second language acquisition: An introductory course*. Routledge, 2013. 28
- [45] F. Gates. *Teachers Know Best. Making Data Work For Teachers and Students*. Bill & Melinda Gates Foundation, 2015. 16
- [46] J. González, R. Wagenaar, et al. *Tuning educational structures in Europe*. University of Deusto Final report. Phase one. Bilbao, 2003. 3, 27, 28
- [47] 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
- [48] 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

- [49] 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
- [50] A. R. Hevner. Interview with Alan R. Hevner on “Design Science”. *Business & Information Systems Engineering (BISE)*, 1(1):126–129, 2009. 6
- [51] 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
- [52] 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. 97
- [53] 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
- [54] A. Huhta. Diagnostic and formative assessment. *The handbook of educational linguistics*, 33:469–483, 2008. 29
- [55] 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
- [56] 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
- [57] M. Ladrón de Guevara Cervera, J. Hincapié, J. Jackman, O. Herrera, and C. V. Caballero Uribe. Revisión por pares:¿ qué es y para qué sirve? *Revista Salud Uninorte*, 24(2):258–272, 2008. 103

BIBLIOGRAFÍA

- [58] 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
- [59] D. Laurillard. Teaching as a design science. *Building pedagogical patterns for learning and technology*, 2012. 5, 7
- [60] C. S. E. Lidz. *Dynamic assessment: An interactional approach to evaluating learning potential*. Guilford Press, 1987. 30
- [61] H. Malehorn. Ten measures better than grading. *The Clearing House*, 67(6):323–324, 1994. 37, 62
- [62] 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
- [63] 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
- [64] A. Masip-Álvarez, C. Hervada-Sala, T. Pàmies-Gómez, A. Arias-Pujol, C. Jaen-Fernandez, C. Rodriguez-Sorigue, D. Romero-Duran, F. Nejari-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
- [65] 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
- [66] V. Midoro. Guidelines on adaptation of the UNESCO ICT competency framework for teachers. 2013. 4

- [67] 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. 97
- [68] 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
- [69] B. J. Oates. *Researching information systems and computing*. Sage, 2005. 12
- [70] 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
- [71] 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. 79
- [72] M. Palomo-Duarte, J. M. Dodero, I. Medina-Bulo, E. J. Rodríguez-Posada, and I. Ruiz-Rube. Assessment of collaborative learning experiences by graphical analysis of wiki contributions. *Interactive Learning Environments*, 22(4):444–466, 2014. 79
- [73] Y. Park and I.-H. Jo. Development of the learning analytics dashboard to support students’ learning performance. *Journal of Universal Computer Science*, 21(1):110–133, 2015. 90
- [74] 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

BIBLIOGRAFÍA

- [75] 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
- [76] 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
- [77] QAA. *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education*. The Quality Assurance Agency for Higher Education, 2006. 30
- [78] 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
- [79] 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, 90
- [80] C. Redecker and y. Johannessen. Changing Assessment — Towards a New Assessment Paradigm Using ICT. *European Journal of Education*, 48(1):79–96, 2013. 25
- [81] P. Reimann. Design-based research. In *Methodological Choice and Design*, pages 37–50. Springer, 2011. 7
- [82] 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

- [83] 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
- [84] 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
- [85] 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
- [86] 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
- [87] 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
- [88] 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
- [89] 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
- [90] 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

BIBLIOGRAFÍA

- [91] 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
- [92] G. Trentin. Using a wiki to evaluate individual contribution to a collaborative learning project. *Journal of Computer Assisted Learning*, 25(1):43–55, 2009. 79
- [93] 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
- [94] R. W. Tyler, R. M. Gagné, M. Scriven, et al. *Perspectives of curriculum evaluation*, volume 1. Rand McNally Chicago, 1967. 29
- [95] A. van Deursen, P. Klint, and J. Visser. Domain-specific languages: an annotated bibliography. *ACM SIGPLAN Notices*, 35(6):26–36, June 2000. 90
- [96] 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
- [97] 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
- [98] 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
- [99] J. S. Wiggins. *Paradigms of personality assessment*. Guilford Press, 2003. 33

BIBLIOGRAFÍA

- [100] 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
- [101] 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 18 de enero
de 2016

This page is intentionally left blank