

Evaluación entre iguales en la enseñanza de Arquitectura

Aplicación en la enseñanza de Sistemas de Información Geográfica

Peer-review in Architecture education

Application in Geographic Information Systems learning

Pilar Garcia-Almirall, Ernest Redondo, Francesc Valls

Escuela Técnica Superior de Arquitectura de Barcelona

UPC - BarcelonaTech

Barcelona, España

{pilar.garcia-almirall, ernesto.redondo, francesc.valls}@upc.edu

David Fonseca

La Salle, Campus Barcelona

Universitat Ramon Llull (URL)

Barcelona, España

fonsi@salle.url.edu

Resumen — Este artículo describe una experiencia docente en el ámbito de la enseñanza de los Sistemas de Información Geográfica (SIG) en Arquitectura con un doble objetivo: (1) implementar un ejercicio de evaluación entre iguales, con el objetivo de fomentar el pensamiento crítico a través la evaluación de los trabajos de los compañeros y compañeras de curso, y (2) explorar la capacidad del grupo para evaluar los trabajos con el soporte de una rúbrica. Además del planteamiento de la experiencia, se exponen los datos preliminares del ensayo que muestra una importante coincidencia entre la evaluación propia y las evaluaciones ajenas.

Palabras Clave - evaluación entre iguales; arquitectura; SIG; rúbrica; educación.

Abstract — This paper describes an educational experience in the field of Geographic Information Systems (GIS) teaching in architecture with a double objective: (1) implement a peer-review exercise to promote critical thinking through the evaluation of the work of fellow students, and (2) explore the capacity of the group to assess the exercise with the support of a rubric. In addition to the outline of the experience, preliminary data showing a significant coincidence between self and peer-evaluation is discussed.

Keywords - peer-review; architecture; GIS; rubric; education.

I. INTRODUCCIÓN

El uso de la evaluación entre iguales (peer-review) se utiliza en ciencia para garantizar la revisión objetiva de la producción científica, generalmente de manera anónima para una de las partes (blind peer-review) o las dos partes (double-blind peer-review), tanto evaluador como evaluado.

Esta relación entre iguales no solo se circunscribe al ámbito de la ciencia, sino que en Internet el sistema de reputación implementado en sitios como Stack Exchange, la votación de los elementos más populares en sitios como Reddit o la moderación colaborativa como la que encontramos en Wikipedia permiten resultados comparables o superiores a los realizados por profesionales [1], dónde el trabajo de los

individuos resulta en productos de los que se beneficia toda la comunidad.

El doble objetivo de la experiencia docente explicada en este artículo es (1) implementar una evaluación entre iguales para la enseñanza de Sistemas de Información Geográfica (SIG) en Arquitectura, con el objetivo de fomentar el pensamiento crítico durante la evaluación de los trabajos de los compañeros y compañeras de curso, y (2) explorar la capacidad del grupo para evaluar los trabajos con el soporte de una rúbrica.

A. Evaluación entre iguales (peer-review) en educación

El uso de evaluación entre iguales ha tenido un importante empuje con la llegada de los MOOCs (Massive Open Online Courses), dónde por sus características su uso es casi obligado para evaluar trabajos en los que no es posible la evaluación automática, como por ejemplo en la evaluación de preguntas abiertas o trabajos concretos cuando no es posible reducir la evaluación a un cuestionario de respuesta múltiple. En la literatura encontramos referencias que demuestran que en estos cursos la calificación otorgada por el propio participante a sí mismo es muy parecida a la otorgada por los otros participantes [2], especialmente cuando se tiene el soporte de una rúbrica.

En cuanto a la relación entre la nota asignada por el profesorado comparada con la asignada por la evaluación entre iguales, el estudio comparativo de múltiples experiencias de autoevaluación y evaluación entre iguales [3] concluye que la correlación aumenta cuando los criterios a evaluar son más claros.

En este sentido, la aplicación de rubricas [4] en educación universitaria tiene la capacidad de mejorar el rendimiento académico [5], [6], cuando se aplica en evaluación entre iguales [7], alineando las calificaciones con las del profesorado y fomentando el pensamiento crítico [8].

El uso de estas técnicas en la educación universitaria [9] se enmarca dentro del constructivismo y ofrece ventajas para los participantes en su papel como evaluadores (fomentando el

pensamiento crítico) y como evaluados (mejorando la calidad de su trabajo por el hecho de saber que serán evaluados).

II. MEJORAS DOCENTES

La presente metodología parte de una experiencia previa dónde también se utilizó la evaluación entre iguales. Esta experiencia piloto permitió identificar tres elementos a mejorar, que se han subsanado en la presente edición:

- La necesidad de disponer de una rúbrica
- La inconveniencia de presentar los resultados de la evaluación en un documento
- La formación de los emparejamientos de manera no aleatoria

El primer tercio de las clases se planteó como un taller autocontenido. Una vez finalizado, se pidió a los alumnos que entregaran un trabajo para ser evaluado, consistente en la descripción de los pasos seguidos para llegar a los resultados junto con la descripción de la metodología seguida en el desarrollo del taller.

El objetivo era que el estudiantado pudiera conocer los criterios de evaluación antes de la entrega final de tres maneras distintas: (a) con la evaluación del taller por parte del profesorado, (b) evaluando su propio trabajo, y (c) evaluando el trabajo de dos compañeros. Las tres evaluaciones de los alumnos se hicieron con la misma rúbrica con la que fueron evaluados por el profesorado.

A. Uso de rúbricas

En la experiencia piloto los alumnos dispusieron de unos criterios de evaluación pero a diferencia de la experiencia aquí explicada no estaba formalmente descrita. Ello conllevó una dificultad en la evaluación por parte del alumnado que pudo detectarse en conversaciones informales con ellos. Esta limitación llevó a implementar el uso de una rúbrica en la presente edición. El uso de rúbricas tiene ventajas en dos niveles: (a) ventajas para los alumnos, y (b) ventajas para los profesores.

La principal ventaja para los alumnos es que mediante la rúbrica son conocedores desde el primer día del curso de la producción que se espera de ellos, así como los criterios de evaluación que les serán aplicados. Esta búsqueda de la objetividad en los criterios permite esquivar uno de los mayores problemas en la relación profesorado-alumnado: la falta de transparencia y objetividad en la calificación de los resultados.

Esta transparencia también es una ventaja para el profesorado, puesto que disminuye la posibilidad de conflicto con el alumno cuando aparece un desacuerdo en la evaluación de los resultados. Adicionalmente, el uso de rúbricas permite mantener un criterio unificado durante distintas ediciones de los cursos a lo largo del tiempo, facilitando la incorporación de nuevo profesorado al equipo docente.

Finalmente, en la presente experiencia permitió a los participantes calificar a otros compañeros con un criterio único, y utilizar este mismo criterio para evaluarse a sí mismos. Este mismo criterio es el que siguió el profesorado.

B. Evaluación online

La segunda limitación fue presentar la evaluación en forma de documento. La redacción de este informe supuso una dificultad para los participantes en la prueba piloto puesto que les suponía una importante carga de trabajo, que se ha intentado subsanar en la presente edición mediante la evaluación mediante un cuestionario online.

La plataforma utilizada fue Google Forms, con la cual los alumnos ya estaban familiarizados porque es la misma que utilizan para responder las encuestas. Esta plataforma tiene la ventaja que los alumnos pueden responder el cuestionario de evaluación sin necesidad de escribir un documento y desde cualquier lugar o dispositivo conectado a Internet. La plataforma simplifica también la recogida de datos por parte del profesorado ya que los resultados se almacenan un documento en la nube.

El cuestionario se organizó en tres partes: (a) nombre del evaluador y número del trabajo evaluado —en el caso de autoevaluación el código era 0—, (b) valoración de seis aspectos esenciales del trabajo, y (c) valoración general.

También se incluyeron dos apartados adicionales: (a) un apartado para escribir comentarios adicionales de estilo libre sobre los trabajos realizados, y (b) un apartado que permitía retractar una valoración y sustituirla por otra, indicando el motivo del cambio.

El estudiantado tenía que evaluar dos trabajos de compañeros, que se descargaban de la intranet docente y el suyo propio, de manera que debían rellenar un total de tres cuestionarios. Para ello se les hizo entregar los trabajos en formato electrónico, del cual se eliminó cualquier dato identificativo, tanto en el contenido como en los metadatos.

A continuación se colgó en la intranet el conjunto de trabajos, con un nombre de archivo que correspondía a un número aleatoriamente asignado, juntamente con una tabla con sus DNI (Documento Nacional de Identidad) y los números de los dos trabajos que les correspondía evaluar.

C. Emparejamiento

La tercera dificultad parte de la manera como se realizaron los emparejamientos en la primera experiencia: en primer lugar el profesor corrigió los trabajos a evaluar y una vez evaluados se separó a los alumnos en dos grupos que correspondían a los que estaban por encima de la mediana y los que estaban debajo, para aumentar el contraste entre corrector y corregido. La estrategia es que los peor calificados pudieran encontrar motivador ver los trabajos de los mejor calificados e intentar mejorar.

En esta edición los emparejamientos se hicieron de manera aleatoria para no introducir sesgos en el análisis, puesto que se detectó que este contraste tenía consecuencias negativas tanto para los mejores cualificados como para los peor calificados, en el primer grupo porque les inducía a pensar que no era necesario demasiado esfuerzo y en el segundo caso porque se podía producir una cierta sensación de frustración.

III. EXPERIENCIA DOCENTE

A. Definición de la rúbrica

La valoración de los trabajos se hizo en dos niveles: (a) se valoraron seis aspectos esenciales siguiendo unos criterios objetivos, (b) la calidad subjetiva en dos preguntas; además se facilitó un espacio para dejar comentarios adicionales sobre la valoración del trabajo.

Las seis categorías evaluadas se detallan en la tabla que aparece a continuación (Tabla 1), dónde se procuró que los criterios fueran tan objetivos como fuera posible de manera que pudieran ser más fácilmente evaluables:

TABLA I. CRITERIOS DE EVALUACIÓN DE LA RÚBRICA

| | Insatisfactorio | Cumple mínimos | Satisfactorio | Excepcional |
|------------------|---|--|--|--|
| Completación | No desarrolla alguno de los procesos explicados en clase | Describe todos los procesos explicados en clase de manera esquemática | Describe los procesos desarrollados en clase detalladamente | Describe críticamente los procesos desarrollados en clase detalladamente y aporta como mínimo una alternativa metodológica |
| Presentación | No hay ninguna coherencia en la presentación | Hay alguna incoherencia en la presentación (tipos de letra, espaciado, márgenes, etc.) | La presentación presenta un estilo unificado | Hay un esfuerzo en facilitar la comprensión a través de la presentación, utilizando esquemas, tablas, enumeraciones, etc. |
| Estructura | No hay estructuración en apartados y el orden de la explicación es incoherente | No hay estructuración en apartados | Se estructura lógicamente en apartados | Se estructura en apartados y cada apartado se estructura con una introducción, procesos seguidos y conclusiones |
| Salidas gráficas | No hay salidas gráficas de los resultados finales | Hay salidas gráficas de los resultados finales únicamente | Hay salidas gráficas de los resultados finales y de algunos intermedios | Además se incluyen dos o más de los elementos siguientes: pie de imagen, leyenda, flecha norte o escala gráfica |
| Conclusiones | No aporta ninguna conclusión acerca de los resultados | Aporta conclusiones en algunos de los resultados | Aporta conclusiones para todos los resultados | Aporta conclusiones razonadas para todos los resultados |
| Redactado | Aparecen sistemáticamente más de cinco errores de ortografía o sintaxis por párrafo | Se encuentran de media entre cuatro y dos errores de ortografía o sintaxis por párrafo | Existen alrededor de un error de ortografía o sintaxis por párrafo o menos | No se aprecia prácticamente ningún error de ortografía o sintaxis |

Las dos preguntas sobre la calidad subjetiva del trabajo, evaluadas en una escala de 0 a 10, fueron: (a) calidad general del trabajo, y (b) nivel de comprensión adquirido. El objetivo de estas preguntas fue comparar la valoración objetiva con la subjetiva para comprobar su grado de concordancia con cada una de las seis categorías.

B. Emparejamiento de los participantes

En una primera experiencia piloto, los participantes se emparejaron con un criterio de máximo contraste: primero los trabajos fueron corregidos por el profesorado y se les asignó una calificación, en segundo lugar los emparejamientos se realizaron de manera que (a) las diferencias entre participantes fueran máximas y (b) la variación dentro de las diferencias fuera mínima.

En la experiencia piloto, los emparejamientos se hicieron por desfase (Código 1) ordenando de menor a mayor los resultados de la evaluación y asignándoles el trabajo del participante en la posición N/2 posiciones para ser evaluado (Código 2), de manera que el de menor calificación se emparejara con el participante cuya calificación que se encontraba en la mediana, el segundo valor con el siguiente y así sucesivamente (Fig. 1).

```
shift <- function(x, offset){
  wrap <- length(x)
  return(1 + ( x + offset - 1 ) %% wrap)
}
```

Código 1: Script en R para decalar los emparejamientos

```
students.sorted <- 1:6
N <- length(students.sorted)
peer.contrast <- shift(students.sorted, N%%2)
```

Código 2: Script en R para el emparejamiento con máximo contraste de seis participantes, utilizando el script 1

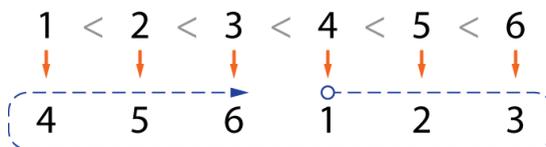


Figura 1: Emparejamientos en la prueba piloto

En esta ocasión, el emparejamiento de los participantes se hizo de forma aleatoria, debido en primer lugar a los inconvenientes que se detectaron en la experiencia piloto cuando se emparejaron los participantes con el criterio de máximo contraste y en segundo lugar para no distorsionar los datos con sesgos introducidos por la selección. Como ventaja añadida, esta metodología no requería que los trabajos fueran corregidos por el profesorado como paso previo al reparto, hecho muy importante en una asignatura intensiva por escaso tiempo disponible entre clases.

La manera de emparejar los participantes fue ordenar los participantes de forma aleatoria y emparejar los correctores con el resultado de los participantes cuyo orden estuviera desfasado en una fracción del número total de trabajos a evaluar (Código

3). De esta manera se evita que haya reciprocidad entre las correcciones entre participantes, es decir, si A corrige a E y C, éstos últimos no corrijan asimismo a A, siendo esta regla válida para todos los correctores y corregidos (Fig. 2).

```

pairing <- function(x, groups){
  N <- length(x)
  if(groups >= N) {stop("too many groups")}
  m <- x
  for(i in 1:groups){
    order <- shift(1:N, i)
    p <- x[order]
    m <- cbind(m, p, deparse.level = 0)
  }
  return(m)
}

students.rnd <- sample(1:6)
pairing.matrix <- pairing(students.rnd, 2)

```

Código 3: Script en R para obtener la matriz con el emparejamiento aleatorio de seis participantes sin reciprocidad, utilizando el script 1

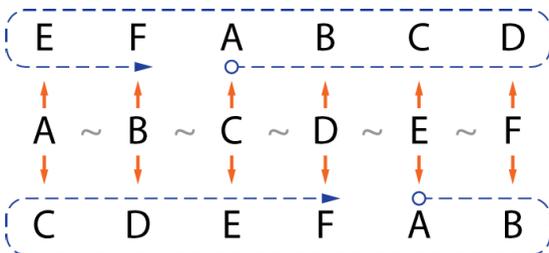


Figura 2: Emparejamientos aleatorios entre un evaluador y dos trabajos de otros compañeros

El resultado es una red donde cada participante evalúa a un total de n participantes, y es evaluado asimismo por n compañeros o compañeras (Fig. 3), es decir el “in-degree” y el “out-degree” son el mismo e igual número de evaluaciones que deben realizar.

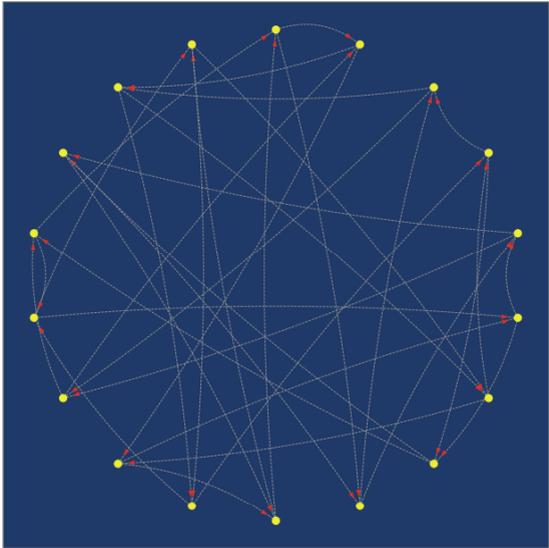


Figura 3: Grafo orientado de las relaciones entre participantes (18 estudiantes y 2 evaluaciones cada uno)

IV. RESULTADOS PRELIMINARES

Se ha realizado una exploración preliminar de los resultados de la evaluación por pares recogidos con Google Forms. Esta primera aproximación se ha centrado en visualizar de manera agregada la diferencia entre la evaluación ajena y la autoevaluación para descubrir si se aprecian diferencias entre los resultados de una y otra. Esta exploración de los resultados se ha hecho en dos bloques que se detallan a continuación:

- Evaluación subjetiva (general y respecto a la comprensión)
- Evaluación mediante rúbrica (seis apartados con cuatro grados posibles cada uno)

A. Evaluación subjetiva

La evaluación subjetiva consistió en evaluar en una escala de 0 a 10 dos aspectos subjetivos del trabajo: (a) la calidad general y (b) la comprensión de los temas de la asignatura demostrada a través del trabajo, analizando sus distribuciones (Fig. 4). Para la evaluación ajena se estableció un peso relativo de un 50% respecto a la propia debido a que dos evaluaciones ajenas que se comparaban con una única propia.

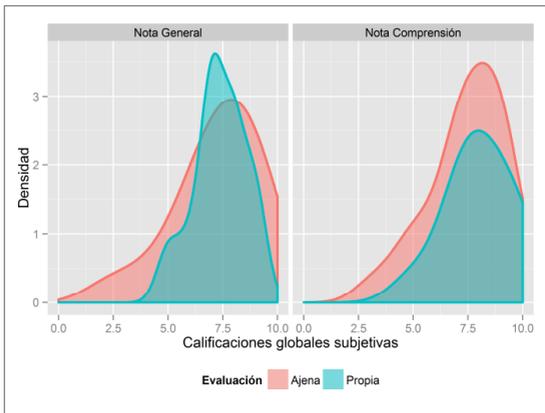


Figura 4: Distribuciones de las dos respuestas subjetivas (respuestas propias y ajenas)

La evaluación de la calidad general del trabajo (Tabla 1) mostró escasa diferencia a nivel agregado entre las medias de la evaluación ajena y la valoración propia (0.1 puntos en una escala de 0 a 10), mientras que su desviación estándar era mayor en el caso de la valoración ajena (0.77 puntos).

TABLA II. RESULTADOS DE LA EVALUACIÓN SUBJETIVA GENERAL EN UNA ESCALA DE 0 A 10

| Evaluación | Media | Mediana | Desv. E. |
|------------|-------|---------|----------|
| Ajena | 7.26 | 8 | 1.98 |
| Propia | 7.36 | 7 | 1.21 |

En el caso de la evaluación de la comprensión de los temas de la asignatura (Tabla 3), la diferencia entre las medias fue ligeramente mayor (0.25 puntos mayor en la autoevaluación) y en este caso las desviaciones estándar son más parecidas (0.25 puntos mayor en la evaluación ajena).

TABLA III. RESULTADOS DE LA EVALUACIÓN SUBJETIVA DE LA COMPRENSIÓN EN UNA ESCALA DE 0 A 10

| Evaluación | Media | Mediana | Desv. E. |
|------------|-------|---------|----------|
| Ajena | 7.45 | 8 | 1.73 |
| Propia | 8.00 | 8 | 1.48 |

Para investigar en más detalle la diferencia entre la evaluación subjetiva en la autoevaluación respecto a la evaluación por iguales se representaron los resultados en una gráfica con las características siguientes (Fig. 5): (a) se representó la frecuencia de las calificaciones utilizando histogramas, (b) se representaron los histogramas en una retícula con la misma escala, con las dos preguntas en columnas y el tipo de evaluación (propia o ajena) en las filas, con colores distintos, y (c) se superpuso la correspondiente curva de densidad a cada histograma.

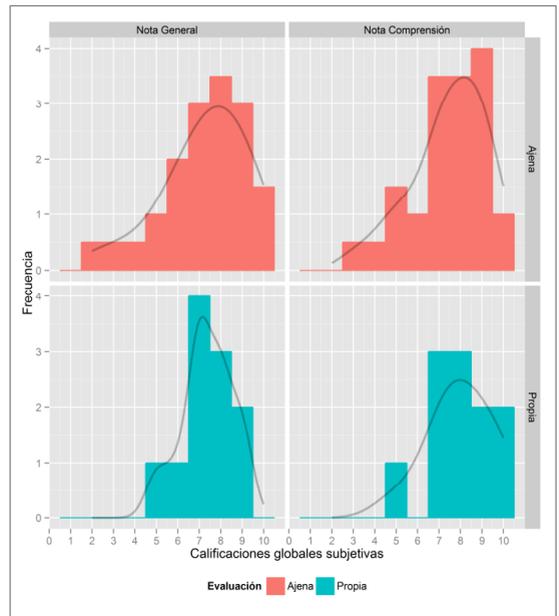


Figura 5: Relación entre la evaluación propia y ajena en la valoración subjetiva del trabajo

Estos resultados preliminares a nivel agregado permitieron comprobar varios factores que deberían ser analizados en profundidad: (a) la evaluación propia y la ajena no son apreciablemente diferentes, (b) los estudiantes se sobrevaloran ligeramente a nivel general y se minusvaloran ligeramente a nivel de comprensión, sin embargo esto es debido a la presencia de más notas altas en la valoración propia, puesto que el grueso de la distribución muestra el efecto contrario, y (c) la valoración de la comprensión es ligeramente mayor que la general en los dos casos.

B. Evaluación mediante rúbrica

La evaluación mediante rúbrica pretendía establecer unas pautas para evaluar seis aspectos del trabajo presentado por el estudiantado (Tabla 1) en cuatro niveles categóricos (insatisfactorio, cumple mínimos, satisfactorio y excepcional).

Se analizó la distribución de los seis aspectos (compleción, presentación, estructura, salidas gráficas, conclusiones y redactado) para la evaluación ajena y propia (Fig. 6) en una retícula compartiendo la misma escala. Los resultados agregados permitieron observar los siguientes aspectos: (a) las distribuciones son parecidas excepto para el caso del redactado, donde los participantes valoran mucho mejor los trabajos ajenos que los propios, (b) los participantes sobrevaloran ligeramente su presentación y subvaloran ligeramente sus conclusiones, (c) las valoraciones propias son generalmente inferiores a las ajenas, y (d) las salidas gráficas son el aspecto mejor valorado, mientras que las conclusiones son el menos valorado.

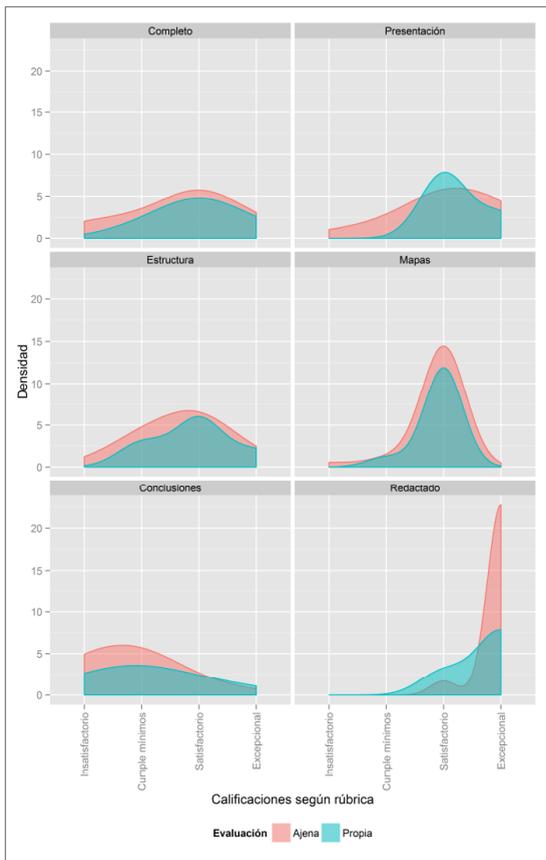


Figura 6: Distribuciones de los seis aspectos evaluados por la rúbrica (autoevaluación en verde y evaluación de otros participantes en rojo)

V. CONCLUSIONES

Las valoraciones preliminares –realizadas inicialmente a nivel agregado– muestran unas tendencias que deberían ser analizadas con más detalle en el futuro. La previsión es que este análisis pormenorizado se realice de dos maneras distintas:

- Realizando más experimentos (en asignaturas de grado y posgrado) que contradigan o confirmen los hallazgos
- Analizando los resultados obtenidos en más detalle

A. Aspectos a analizar

Si se observan los datos de este estudio en mayor detalle se aprecian detalles que escapan al análisis agregado. Por ejemplo, si se representan los datos en un grafo (Fig. 7) se aprecia que el participante P2 no ha realizado ninguna evaluación y que la valoración de su trabajo por otros compañeros ha sido negativa, sugiriendo que se trata de dos aspectos que pueden estar relacionados.

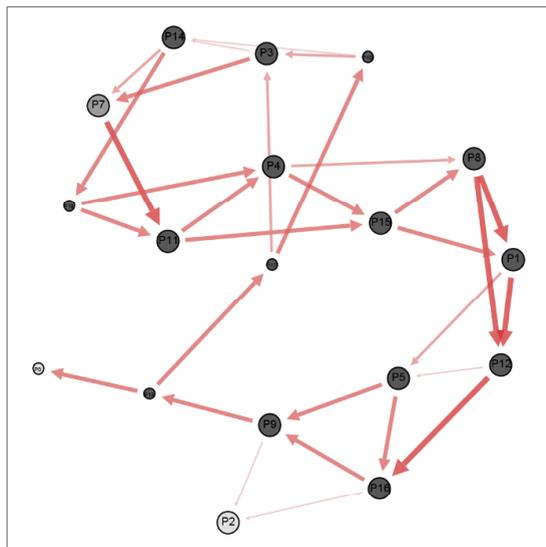


Figura 7: Grafo con las respuestas de la valoración mediante rúbrica del grado de completación del trabajo (tamaño del nodo como número de valoraciones recibidas, color del nodo como número de valoraciones hechas, grueso y color de la flecha como nota de la valoración, flecha de valorador a valorado)

Se propone analizar distintos aspectos en más detalle (a nivel de muestra en lugar de agregadamente), utilizando los datos obtenidos y los de futuras ediciones de la asignatura, modelando estadísticamente y visualizando las relaciones entre distintos aspectos: (a) analizar la relación entre la valoración objetiva (con rúbricas) y subjetiva, (b) medir la objetividad de la autoevaluación en relación con las evaluaciones del resto de compañeros, y (c) comparar la valoración de los participantes con la del profesorado.

Adicionalmente, se recogerá la valoración del estudiantado de la experiencia mediante encuestas cuantitativas [10] y mediante BLA (Bipolar Laddering Assessment) para obtener datos cualitativos [11].

AGRADECIMIENTOS

La presente investigación se ha hecho en el contexto del proyecto EDU2012-37247, E-LEARNING 3.0 EN LA DOCENCIA DE LA ARQUITECTURA. CASOS DE ESTUDIO DE INVESTIGACION EDUCATIVA PARA UN FUTURO INMEDIATO.

REFERENCIAS

- [1] J. Giles, "Internet encyclopaedias go head to head," *Nature*, vol. 438, no. 7070, pp. 900–901, Dec. 2005.
- [2] C. Kulkarni, K. P. Wei, H. Le, D. Chia, K. Papadopoulos, J. Cheng, D. Koller, and S. R. Klemmer, "Peer and Self Assessment in Massive Online Classes," *ACM Trans Comput-Hum Interact*, vol. 20, no. 6, pp. 33:1–33:31, Dec. 2013.
- [3] N. Falchikov and J. Goldfinch, "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks," *Rev. Educ. Res.*, vol. 70, no. 3, pp. 287–322, Sep. 2000.
- [4] D. D. Stevens, A. J. Levi, and B. E. Walvoord, *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback*.

and *Promote Student Learning*, 2 edition. Sterling, Va: Stylus Publishing, 2012.

- [5] Y. M. Reddy and H. Andrade, "A review of rubric use in higher education," *Assess. Eval. High. Educ.*, vol. 35, no. 4, pp. 435–448, Jul. 2010.
- [6] A. Jonsson and G. Svingby, "The use of scoring rubrics: Reliability, validity and educational consequences," *Educ. Res. Rev.*, vol. 2, no. 2, pp. 130–144, 2007.
- [7] J. C. Hafner and P. M. Hafner, "Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating," *Int. J. Sci. Educ.*, vol. 25, no. 12, pp. 1509–1528, 2003.
- [8] H. G. Andrade, "Using rubrics to promote thinking and learning," *Educ. Leadersh.*, vol. 57, no. 5, pp. 13–18, 2000.
- [9] K. Topping, "Peer Assessment Between Students in Colleges and Universities," *Rev. Educ. Res.*, vol. 68, no. 3, pp. 249–276, Sep. 1998.
- [10] P. Garcia-Almirall, E. Redondo Domínguez, F. Valls Dalmau, and J. M. Corso Sarmiento, "Experiencia docente en la enseñanza de Sistemas de Información Geográfica en Arquitectura," in *9th Iberian Conference on Information Systems and Technologies*, Barcelona, 2014, pp. 407–412.
- [11] M. Pifarré and O. Tomico, "Bipolar Laddering (BLA): A Participatory Subjective Exploration Method on User Experience," in *Proceedings of the 2007 Conference on Designing for User eXperiences*, New York, NY, USA, 2007, pp. 2:2–2:13.