

# Integrando Resultados Experimentales sobre Medidas para Diagramas de Clases UML: Un Estudio de Meta-Análisis

M. Esperanza Manso<sup>1</sup>, José A. Cruz-Lemus<sup>2</sup>, Marcela Genero<sup>2</sup>, Mario Piattini<sup>2</sup>

<sup>1</sup>Grupo de investigación GIRO, Departamento de Informática, Universidad de Valladolid.  
Campus Miguel Delibes, E.T.I.C., 47011, Valladolid, España.

manso@infor.uva.es

<sup>2</sup>Grupo de investigación ALARCOS, Departamento de Tecnologías y Sistemas de  
Información. Universidad de Castilla-La Mancha, Paseo de la Universidad, 4. 13071 Ciudad  
Real, España.

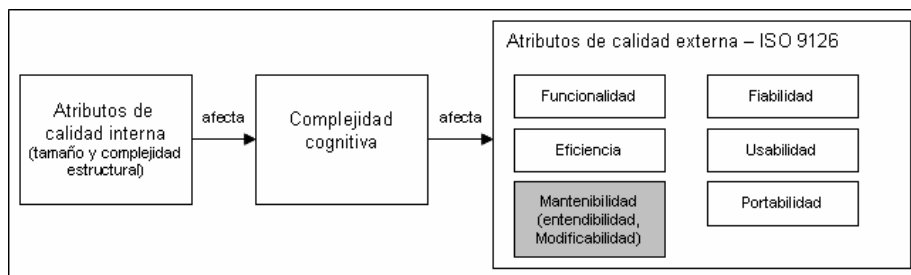
{JoseAntonio.Cruz, Marcela.Genero, Mario.Piattini}@uclm.es

**Resumen.** El objetivo principal de este artículo es realizar un estudio de meta-análisis para integrar los resultados obtenidos en una familia de cinco experimentos controlados realizados en entornos académicos. Estos experimentos se realizaron para validar empíricamente dos hipótesis: por un lado la correlación entre la complejidad estructural y el tamaño de los diagramas de clases UML y su complejidad cognitiva y, por otro, la correlación entre la complejidad cognitiva y la entendibilidad y modificabilidad de los diagramas de clases UML. A través del meta-análisis llegamos a la conclusión de que las medidas relativas a asociaciones y generalizaciones tienen fuerte correlación con la complejidad cognitiva, y la complejidad cognitiva está más correlacionada con la entendibilidad que con la modificabilidad. Estos resultados tienen implicaciones tanto desde el punto de vista del modelado como el de la enseñanza, ya que muestran evidencia empírica sobre cuáles son los constructores UML que tienen más influencia cuando los modeladores tienen que comprender y modificar los diagramas de clases UML. Además, las medidas relativas a asociaciones y generalizaciones pueden ser utilizadas para construir modelos de predicción sobre el esfuerzo de mantenimiento de los diagramas de clases UML.

## 1 Introducción

El Desarrollo de software dirigido por modelos (DSDM) [2] es una propuesta emergente para el desarrollo de software que está cobrando cada vez más interés tanto para la comunidad investigadora como industrial. DSDM considera a los modelos como productos finales más que como un simple medio para producir software. Por lo tanto, en el contexto del DSDM, el mantenimiento se realiza sobre los modelos. Nuestro interés se centra en el mantenimiento de los diagramas de clases UML, específicamente en su entendibilidad y modificabilidad. Los diagramas de clases constituyen la columna vertebral del diseño de un sistema y deben ser suficientemente comprensibles y flexibles para que cambios puedan ser incorporados fácilmente.

Nosotros hemos basado nuestro trabajo en el modelo mostrado en la Figura 1 [6, 20]. Este modelo ha sido utilizado como base de una gran cantidad de investigaciones empíricas sobre propiedades estructurales de artefactos software [10, 31]. En este artículo, suponemos que las propiedades estructurales (como la complejidad estructural y el tamaño) de un diagrama de clases UML afectan a su complejidad cognitiva. La complejidad cognitiva puede ser definida como la carga mental que causa el artefacto en la gente que tiene que tratar con él (por ejemplo modeladores, diseñadores, mantenedores, etc.). Una complejidad cognitiva alta puede hacer que un diagrama de clases tenga entendibilidad y modificabilidad reducidas y esto, consecuentemente, podrá afectar a su mantenibilidad.



**Fig. 1.** Relación entre propiedades estructurales, complejidad cognitiva y atributos de calidad externos, basados en [6, 20].

El principal motivo de este artículo es la validación del modelo presentado en la Figura 1. Para ello formulamos dos hipótesis basadas en cada una de las flechas de la figura:

1. El tamaño y la complejidad estructural de los diagramas de clases UML afectan a la complejidad cognitiva,
2. La complejidad cognitiva afecta a la entendibilidad y modificabilidad de los diagramas de clases UML.

Para medir el contenido de cada caja de la Figura 1 hemos definido medidas, que serán introducidas en la Sección 3. Para probar las hipótesis, hemos llevado a cabo 5 experimentos que constituyen una familia de experimentos [12, 13]. El análisis de los datos llevado a cabo en cada experimento individual no nos permitió obtener resultados concluyentes. Esto nos motivó a realizar un estudio de meta-análisis, considerado como un medio adecuado para agregar o integrar los resultados de estudios empíricos [26, 30]. Además, la necesidad del meta-análisis está ganando relevancia en la investigación empírica, como demuestra el hecho de que es un tema recurrente en varios foros relacionados con la Ingeniería de Software Empírica.

El meta-análisis es una herramienta para extraer conclusiones globales de familias de experimentos, y permite estimar el tamaño del efecto global de la familia, medir la exactitud de esta medida y evaluar el grado de significación con respecto a las hipótesis bajo estudio.

El resto del artículo está organizado de la siguiente manera: la Sección 2 presenta el trabajo relacionado; la Sección 3 describe la familia de experimentos. El estudio de meta-análisis se presenta en la Sección 4. Finalmente, la última sección presenta algunas conclusiones y las líneas de trabajo futuras.

## 2 Trabajo Relacionado

Dentro del contexto de la Ingeniería del Software Empírica ha habido un gran interés en realizar estudios empíricos sobre atributos externos de la calidad como la mantenibilidad, la entendibilidad, la modificabilidad, etc. Primeramente, el foco ha estado en el código o artefactos del diseño detallado [1, 5, 11, 16, 25]. Posteriormente, dado el incremento de la relevancia del modelado, el foco cambió a los modelos. La comprensión y modificación de los diagramas UML ha sido el objetivo de una gran cantidad de investigaciones empíricas sobre diagramas UML llevadas a cabo en años recientes [7, 24, 28, 29, 33].

Nuestros primeros trabajos abordaban la influencia de la complejidad estructural y el tamaño de los diagramas de clases UML sobre su entendibilidad y modificabilidad. Un resumen de estos trabajos se muestra en [13]. En todos esos trabajos, sea realizaron varios experimentos controlados, pero el análisis de datos de cada experimento individualmente, en algunos casos obtuvo resultados controvertidos. Por esta razón y debido a la creciente necesidad de investigar los constructores UML que tienen más influencia en la comprensión y modificación de los diagramas de clases UML, decidimos integrar los resultados de experimentos homogéneos a través de un estudio de meta-análisis, que es el principal objetivo del presente trabajo.

## 3 Una Familia de Experimentos

En esta sección comentaremos cinco experimentos controlados, que realizamos teniendo en cuenta algunas recomendaciones propuestas en [21, 36]. La Tabla 1 resume las principales características de contexto de los cinco experimentos.

**Tabla 1.** Características de la familia de experimentos

	#Sujetos	Universidad	Fecha	Curso
E1	72	Universidad de Sevilla (España)	Marzo 2003	4º
R1	28		Marzo 2003	
E2	38	Universidad de Castilla-La Mancha (España)	Abril 2003	3º
R21	23	Universidad de Sannio (Italia)	Junio 2003	4º
R22	71	Universidad de Valladolid (España)	Septiembre 2005	3º

### 3.1 Planificación de los experimentos

En esta subsección definiremos las características comunes a todos los experimentos, consistentes en:

**Preparación.** La familia tiene un doble objetivo, definido como:

- Objetivo 1: analizar la complejidad estructural y el tamaño de los diagramas de clases UML con respecto a su relación con la complejidad cognitiva desde el punto de vista de los modeladores y diseñadores de software en un contexto académico.

- Objetivo 2: analizar la complejidad cognitiva de los diagramas de clases UML con respecto a su relación con la entendibilidad y modificabilidad desde el punto de vista de modeladores o diseñadores software en un contexto académico.

**Definición del contexto:** en estos estudios hemos utilizado estudiantes como sujetos experimentales. Las tareas a realizar no requerían altos niveles de experiencia industrial, así que creímos que estos sujetos podían considerarse apropiados, como se apunta en varios trabajos [3, 19]. En resumen, trabajar con estudiantes implica una serie de ventajas, como el hecho de que el conocimiento previo es bastante homogéneo, la disponibilidad de un elevado número de sujetos, y la posibilidad de probar diseños experimentales e hipótesis iniciales [34]. Una ventaja más es el uso de principiantes como sujetos en experimentos en entendibilidad es que la complejidad cognitiva del sujeto bajo estudio no es tan alta como la de sujetos con experiencia.

**Material:** los materiales experimentales consisten en un conjunto de diagramas de clases UML que cubren un amplio rango de valores las medidas [14, 15]. Así, obtuvimos tres tipos de diagramas: difíciles de mantener (D), fáciles de mantener (F) y de una dificultad mediana (M). Algunos fueron diseñados específicamente para los experimentos y otros se obtuvieron de aplicaciones reales. Cada diagrama tiene documentación añadida que contiene, además de otras cosas, cuatro tareas de comprensión y cuatro de modificación.

### 3.2 Conducción de los experimentos individuales

Las variables consideradas para medir la complejidad estructural y el tamaño fueron un conjunto de 11 medidas presentadas en la Tabla 5 (Apéndice A). La medida *CompSub* es la percepción subjetiva de los sujetos sobre la complejidad de los diagramas con los que trabajaron durante la tarea experimental. Nosotros consideramos *CompSub* como una medida de la complejidad cognitiva. Los posibles valores de esta variable son: Muy sencillo, Moderadamente sencillo, Medio, Moderadamente complejo y Muy complejo. Para medir la entendibilidad y modificabilidad de los diagramas consideramos el tiempo (en segundos) utilizado por cada sujeto para completar las preguntas de entendibilidad y modificabilidad. Hemos llamado a estas medidas Tiempo de entendibilidad y modificabilidad.

Utilizamos un diseño balanceado inter-sujetos, así cada sujeto trabajó con un único diagrama. Los diagramas se asignaron aleatoriamente y cada diagrama fue asignado al mismo número de sujetos.

Formulamos las siguientes hipótesis, que se derivan de los objetivos de la familia, previamente presentados:

- $H_{0,1}$ : la complejidad estructural y el tamaño de los diagramas de clases UML no está correlacionado con la complejidad cognitiva.  $H_{1,1}:\neg H_{0,1}$
- $H_{0,2}$ : La complejidad cognitiva de los diagramas de clases UML no está correlacionada con su entendibilidad y modificabilidad.  $H_{1,2}:\neg H_{0,2}$

Todos los experimentos fueron supervisados y limitados en el tiempo. Se pueden encontrar más detalles en [12, 13].

Utilizamos SPSS [35] para realizar todos los análisis estadísticos y la herramienta Comprehensive Meta-Analysis v.2 [4] para realizar el meta-análisis.

### 3.3 Experimento 1 (E1) y réplica (R1)

Comprobando las hipótesis a través del coeficiente de correlación de Spearman, hemos obtenido las siguientes conclusiones relacionadas con los dos objetivos de la familia de experimentos:

- Objetivo 1: la complejidad estructural y la complejidad cognitiva presentan una correlación positiva significativa para todas las medidas, con la excepción de NM, NGEN y MAXDIT en R1.
- Objetivo 2: la complejidad cognitiva parece estar correlacionada positivamente con el esfuerzo necesario para comprender los diagramas de clases UML, pero los resultados son significativos sólo para E1. Al mismo tiempo, no hay correlación con el esfuerzo necesario para modificar los diagramas. Una posible explicación para esto puede ser que los sujetos basen su percepción en la dificultad de la primera prueba que realizan, que en este caso es la de comprensión.

### 3.4 Experimento 2 (E2) y sus réplicas (R21 y R22)

En estos estudios, los objetivos y variables fueron los mismos que en los descritos previamente, pero los diagramas utilizados fueron diferentes, y el contexto y diseño fueron mejorados. Se puede encontrar información más detallada en [13].

De nuevo, hemos utilizado a sujetos elegidos por conveniencia, pero en este caso se ha mejorado la selección bloqueando la experiencia de los sujetos. Realizamos un test previo, y con los resultados obtenidos dividimos a los sujetos en dos grupos. Cada diagrama fue asignado al mismo número de sujetos en cada grupo. Se pueden encontrar más detalles de este proceso en [13].

Las medidas de entendibilidad y modificabilidad sólo se incluyeron cuando la realización de las tareas tenía un nivel de calidad mínimo (corrección y completitud). Los sujetos que estuvieron por debajo del 75% en corrección y completitud fueron excluidos del estudio.

Después de comprobar las hipótesis formuladas y relacionándolas con los objetivos de la familia de experimentos, concluimos que:

- Objetivo 1: tenemos resultados favorables que admiten una correlación entre la complejidad estructural y la complejidad cognitiva de los diagramas de clases UML. Muchas de las medidas están significativamente correlacionadas con la complejidad subjetiva en diferentes estudios, especialmente en aquellos relacionados con las jerarquías de herencia (ver Tabla 2).
- Objetivo 2: los resultados también están a favor de la hipótesis que relaciona la complejidad cognitiva con la entendibilidad de los diagramas de clases UML.

**Tabla 2.** Medidas correlacionadas en E1, R21 y R22

<b>Estudio</b>	<b>Medidas correlacionadas significativamente</b>
E2	NC, NAssoc, NGen, NGenH, MaxDIT (5 sobre 11)
R21	Todas excepto NM, NGenH y MaxAgg (8 sobre 11)
R22	Todas excepto NM (10 sobre 11)

## 4 Estudio de Meta-Análisis

El meta-análisis consiste en un conjunto de técnicas estadísticas que nos permiten combinar medidas del tamaño del efecto de los experimentos individuales para obtener un efecto global. Como las medidas pueden proceder de diferentes entornos e incluso no ser homogéneas, es necesario obtener una medida estandarizada de cada una para poder combinarlas obteniendo así el tamaño del efecto global. Existen diferentes estimadores del tamaño del efecto individual, como las diferencias de medias o los coeficientes de correlación, entre otros. A partir de ellos se estima el tamaño del efecto global como un promedio ponderado; en el que el peso que se utiliza generalmente es el tamaño muestral o la desviación estándar [18]. Además de una estimación puntual del tamaño del efecto global se suele añadir información sobre estimación con un intervalo de confianza y un *p-valor* que permita decidir sobre la hipótesis del meta-análisis.

El meta-análisis está cobrando cada vez más interés en la Ingeniería del Software Empírica, hecho demostrado por las numerosas y recientes publicaciones al respecto [8, 17, 23, 27, 32]. El uso del meta-análisis nos permite extraer conclusiones globales, a pesar de que algunas condiciones experimentales pueden ser diferentes. Como hemos mencionado previamente, necesitamos estandarizar los tamaños de los efectos. En este meta-análisis hemos utilizado coeficientes de correlación para a partir de ellos, obtener la métrica *g de Hedges* [18, 22] de esos valores. Para estudios en Ingeniería del Software, podemos clasificar los tamaños de los efectos en pequeño, mediano y grande [22], tal y como se hace en otras áreas como medicina o psicología. que utilizaremos como indicador del grado en que los datos apoyan la hipótesis de interés.

En nuestro caso el meta-análisis se realizó con los cinco estudios empíricos previamente presentados. Las hipótesis de interés, derivadas de los objetivos de este trabajo, podemos enunciarlas así:

1. Cuanto mayor tamaño o complejidad estructural tenga un diagrama de clases UML mayor complejidad cognitiva tendrá.
2. Cuanta mayor complejidad cognitiva tenga un diagrama de clases UML más tiempo necesitaremos para entenderlo o modificarlo.

De acuerdo con ellas los test de hipótesis estadísticos se definen con las hipótesis nulas y alternativas siguientes:

- $H_{0,1}$ : la complejidad estructural y el tamaño de los diagramas de clases UML no están correlacionados positivamente con la complejidad cognitiva.  $H_{1,1}:\neg H_{0,1}$
- $H_{0,2}$ : La complejidad cognitiva de los diagramas de clases UML no está correlacionada positivamente con su entendibilidad y modificabilidad.  $H_{1,2}:\neg H_{0,2}$

### 4.1 Resultados del meta-análisis

Primeramente, realizamos un meta-análisis para cada par *Medida-CompSub*, teniendo en cuenta los test de hipótesis ( $H_{0,i}$ ) planteados anteriormente. En la Tabla 3 presentamos la estimación global de los coeficientes de correlación, con un intervalo

de confianza del 95%, el *p-valor* y el valor de la *g de Hedges*, incluyendo una clasificación del tamaño del efecto como grande (G), mediano (M) o pequeño (P).

**Tabla 3.** Meta-análisis para *Medidas-CompSub*

<b>H0: <math>\rho \leq 0</math></b>	<b>Correlación (<math>\rho</math>) Tamaño del Efecto Global</b>	<b>Límite inferior</b>	<b>Límite superior</b>	<b>p-valor</b>	<b>g de Hedges</b>
NC	0.566	0.464	0.653	0.0000	1.322(G)
NA	0.541	0.435	0.632	0.000	1.219(G)
NM	0.177	0.040	0.307	0.012	0.339(P)
NAssoc	0.566	0.465	0.653	0.000	1.318(G)
NAgg	0.481	0.368	0.581	0.000	1.051(M)
NDep	0.484	0.371	0.584	0.000	1.060(M)
NGen	0.484	0.371	0.584	0.000	1.018 (G)
NGenH	0.422	0.302	0.529	0.000	0.903 (M)
NaggH	0.393	0.270	0.504	0.000	0.814 (M)
MaxDIT	0.492	0.379	0.590	0.000	1.080 (G)
MaxHagg	0.360	0.233	0.474	0.000	0.734 (M)

Los resultados obtenidos están a favor de la existencia de una correlación positiva entre la complejidad cognitiva y las 11 medidas que miden la complejidad estructural y el tamaño de los diagramas de clases UML. De hecho, muchos de los tamaños del efecto son medios o grandes, con la excepción de NM, que tiene un tamaño del efecto pequeño. Las medidas de tamaño que tienen más influencia en la complejidad cognitiva, medida bien con el coeficiente de correlación o con la *g* de Hedges son NC y NA. Entre las medidas de complejidad que tienen más influencia en la complejidad cognitiva se encuentran las relacionadas con las asociaciones (NAssoc) y generalizaciones (NGen y MaxDIT). Podemos concluir que diagramas con gran cantidad de clases y atributos van a incrementar su complejidad cognitiva. Además, los diagramas de clases que se modelan utilizando mucho los mecanismos que proporcionan la herencia, y las asociaciones lo hacen a expensas de aumentar su complejidad cognitiva.

En la Tabla 4 se muestran los resultados del meta-análisis, análogo al anterior pero para estudiar los test de hipótesis ( $H_{0,2}$ ). También en este caso los datos son favorables a la existencia de correlación entre la complejidad cognitiva y la entendibilidad y la modificabilidad de los diagramas de clases UML. Los tamaños del efecto son medios en ambos casos, pero mayor para la entendibilidad (0,684) que para la modificabilidad. (0,368), lo cual representa también una correlación positiva mayor entre la complejidad comitiva y la entendibilidad. Esto significaría que cuanto mayor sea la complejidad cognitiva de un diagrama de clases mayor será el tiempo necesario para entenderlo y modificarlo, pero además, el incremento de tiempo será mayor para entenderlo que para modificarlo.

**Tabla 4.** Meta-análisis de *CompSub*-Tiempo de Entendibilidad y Modificabilidad

H0: $\rho \leq 0$	Correlación ( $\rho$ ) Tamaño del Efecto Global	Límite inferior	Límite superior	p-valor	g de Hedges
Tiempo de Entendibilidad	0.330	0.200	0.449	0.000	0.684 (M)
Tiempo de Modificabilidad	0.186	0.044	0.320	0.011	0.368(M)

## 5 Conclusiones

El objetivo principal de este trabajo es la validación de un modelo teórico que relaciona el tamaño y la complejidad estructural de los diagramas de clases UML y la complejidad cognitiva, con dos de sus atributos externos de calidad: la entendibilidad y la modificabilidad (ver Figura 1). Para este propósito, hemos llevado a cabo un estudio de meta-análisis con los datos obtenidos de una familia de cinco experimentos. Los resultados del meta-análisis están a favor del modelo considerado con respecto a los dos objetivos perseguidos:

- Objetivo 1: las medidas de complejidad estructural están correlacionadas positivamente con la complejidad cognitiva, especialmente aquellas relacionadas con las asociaciones y generalizaciones. En cuanto a las medidas de tamaño, un aumento en el número de clases y de atributos también supone un incremento de la complejidad cognitiva de los diagramas de clases UML.
- Objetivo 2: la complejidad cognitiva influye en el tiempo para entender y el tiempo de modificación de los diagramas de clases UML, especialmente en el primero.

Estos resultados son relevantes porque podrán utilizarse como medios para controlar el nivel de ciertos atributos de calidad de los diagramas de clases UML durante la fase de modelado. Además, tienen implicaciones tanto prácticas como pedagógicas, proporcionando información sobre cuales son los constructores de UML que tienen más influencia sobre el esfuerzo para entender y mantener los diagramas de clases UML. Cuando existan diseños alternativos de un diagrama de clases UML, pueden ser guías de gran utilidad para seleccionar el que minimice estos constructores.

Además, las medidas relacionadas con las asociaciones y generalizaciones pueden ser utilizadas para construir modelos de predicción, que evalúen cuanto se incrementa el tiempo para entender o modificar un diagrama de clases, estudio realizado en [14]. Queda pendiente como trabajo futuro refinar y validar los modelos de predicción obtenidos utilizando los datos de toda la familia de experimentos.

Además seguiremos trabajando para confirmar los resultados obtenidos en el estudio presentado en este artículo, mejorando distintos aspectos relacionados con las amenazas a la validez de las conclusiones, para ello:

- Incrementaremos el tamaño de la muestra de diagramas de clases, dado que la generalización de los resultados depende de la muestra de objetos examinados con

respecto a toda la población de objetos. Este meta-análisis comprende 33 (24+9) diagramas de clases.

- Trabajaremos con sujetos de diferentes campos, preferiblemente profesionales o estudiantes de otras universidades, para poder generalizar los resultados con respecto a la población de sujetos.
- Mejoraremos las tareas de modificación para hacerlas lo más reales posible.
- Investigaremos otras medidas relativas a la complejidad cognitiva.

## Agradecimientos

Esta investigación forma parte del proyecto ESFINGE (TIN2006-15175-C05-05) financiado por el Ministerio de Educación y Ciencia (España), y del proyecto IDONEO (PAC08-0160-6141) financiado por la Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha (España).

## Referencias

1. Arisholm, E. and Sjøberg, D.I.K.: Evaluating the Effect of a Delegated versus Centralized Control Style on the Maintainability of Object-Oriented Software. *IEEE Trans. Soft. Eng.*, 30(8), 521-534 (2004)
2. Atkinson, C. and Kühne, T.: Model Driven Development: a Metamodeling Foundation. *IEEE Trans. Softw. Eng.*, 20, 36-41 (2003)
3. Basili, V., Shull, F. and Lanubile, F.: Building Knowledge through Families of Experiments. *IEEE Trans. Soft. Eng.*, 25, 456-473 (1999)
4. Biostat, Comprehensive Meta-Analysis v2, <http://www.meta-analysis.com/>
5. Briand, L., Bunse, C. And Daly, J.: A Controlled Experiment for Evaluating Quality Guidelines on the Maintainability of Object-Oriented Designs. *IEEE Trans. Soft. Eng.*, 27(6), 513-530 (2001)
6. Briand, L., Morasca, S. and Basili, V.: Defining and Validating Measures for Object-Based High-Level Desing. *IEEE Trans. Soft. Eng.*, 25, 722-743 (1999)
7. Cruz-Lemus, J. A., Genero, M., Manso, M. E. and Piattini, M.: Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams. In: 8th International Conference on Model-Driven Engineering, Languages and Systems (MoDELS 2005), LNCS vol. 3713, 113-125 (2005)
8. Dybå, T., Arisholm, E., Sjøberg, D.I.K., Hannay, J.E. and Shull, F.: Are Two Heads Better than One? On the Effectiveness of Pair Programming. *IEEE Soft.*, 24(6), 10-13 (2007)
9. El-Emam, K., Benlarbi, S., Goel, N. and Rai, S.: The Confounding Effect of Class Size on the Validity of Object-Oriented Metrics. *IEEE Trans. Soft. Eng.*, 27(7), 630-650 (2001)
10. El-Emam, K., and Melo, W.: The Prediction of Faulty Classes Using Object-Oriented Design Metrics. National Research Council of Canada, NRC/ERB1064 (1999)
11. Fioravanti, F. and Nesi, P.: Estimation and Prediction Metrics for Adaptative Maintenance Effort of Object-Oriented Systems. *IEEE Trans. on Softw. Eng.*, 27(12), 1062-1083 (2001)
12. Genero, M., Manso, M.E. and Piattini, M.: Early Indicators of UML Class Diagrams Understandability and Modifiability. *ACM-IEEE Intl. Symp. Emp. Soft. Eng.*, 207-216 (2004)
13. Genero, M., Manso, M.E., Visaggio, A., Canfora, G. and Piattini, M.: Building Measure-Based Prediction Models for UML Class Diagram Maintainability. *Emp. Soft. Eng.*, 12, 517-549 (2007)
14. Genero, M., Piattini, M. and Calero, C.: Early Measures for UML Class Diagrams. *L'Object*, 6(4), 495-515.(2000)

15. Genero, M., Poels, G., Manso, M. E. and Piattini, M.: Defining and Validating Metrics for UML Class Diagrams. In *Metrics for Software Conceptual Models*, Imperial College Press (2005)
16. Harrison, R., Counsell, S. and Nithi, R.: Experimental Assessment of the Effect of Inheritance on the Maintainability of Object-Oriented Systems. *J. Syst. Soft.*, 52, 173-179 (2000)
17. Hayes, W.: Research in Software Engineering: a Case for Meta-Analysis. 6th IEEE International Symp. Softw. Metrics (METRICS'99), 143-151 (1999)
18. Hedges, L. V. and Olkin, I.: *Statistical Methods for Meta-Analysis*, Academia Press (1985)
19. Höst, M., Regnell, B. and Wohlin, C.: Using Students as Subjects - a Comparative Study of Students & Professionals in Lead-Time Impact Assessment. 4th Conf. Emp. Assessm. Eval. Soft.e Eng. (EASE 2000), 201-214 (2000)
20. ISO-IEC, ISO/IEC 9126. *Information Technology - Software Product Quality* (2001)
21. Juristo, N., and Moreno, S.: *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers (2001.)
22. Kampenes, V., Dybå, T., Hannay, J. E. and Sjøberg, D.I.K.: A Systematic Review of Effect Size in Software Engineering Experiments. *Inf. Soft. Tech.*, 49(11-12), 1073-1086 (2007)
23. Laitenberger, O., El-Emam, K. and Harbich, T.: An Internally Replicated Quasy-Experimental Comparison of Checklist and Perspective-based Reading of Code Documents. IESE, 006.99/e, (1999)
24. Lange, C.F.J. and Chaudron, M.R.V.: Interactive Views to Improve the Comprehension of UML Models - An Experimental Validation. 15th IEEE Intl. Conf. Program Comprehen. (ICPC'07), 221-230 (2007)
25. Li, W. and Henry, S.: Object-Oriented Metrics that Predict Maintainability. *J. Syst. Soft.*, 23, 111-122 (1993)
26. Lipsey, M. and Wison, D.: *Practical Meta-Analysis*, Sage (2001)
27. Miller, J. and McDonald, F.: *Statistical Analysis of Two Experimental Studies*. University of Strathclyde, EFoCS-31-98 (1998)
28. Otero, M. C. and Dolado, J. J.: An Empirical Comparison of the Dynamic Modeling in OML and UML. *J. Syst. Softw.*, 77(2), 91-102 (2005)
29. Otero, M. C. and Dolado, J. J.: Evaluation of the Comprehension of the Dynamic Modeling in UML. *Inf. Softw. Tech.*, 46(1), 35-53 (2004)
30. Pickard, M.: *Combining Empirical Results in Software Engineering*. University of Keele, T-R V1 (2004)
31. Poels, G. and Dedene, G.: Measures for Assessing Dynamic Complexity Aspects of Object-Oriented Conceptual Schemes. 19th Intl. Conf. Concept. Model. (ER 2000), 499-512 (2000)
32. Porter, A. and Johnson, M.: Assessing Software Review Measurement: Necessary and Sufficient Properties for Software Measures. *Inf. Soft. Tech.*, 42(1), 35-46 (1997)
33. Purchase, H. C., Colpoys, L., McGill, M. and Carrington, D.: UML Collaboration Diagram Syntax: an Empirical Study of Comprehension. 1st Intl. Ws. Visualizing Softw. Underst. Anal. (VISSOFT'02), 13-22 (2002)
34. SPSS, SPSS 12.0, *Syntax Reference Guide* (2003)
35. Staron, M., Kuzniarz, L. and Wohlin, C.: Empirical Assessment of Using Stereotypes to Improve Comprehension of UML Models: a Set of Experiments. *J. Syst. Softw.*, 79, 727-742 (2006)
36. Wohlin, C., Runeson, P., Hast, M., Ohlsson, M. C., Regnell, B. and Wesslen, A.: *Experimentation in Software Engineering: an Introduction*. Kluwer Academic Publisher (2000)