

# Ingeniería de Software Basada en Evidencias y Revisiones Sistemáticas



## Evaluación de calidad

Evidence-based software engineering and systematic reviews –  
Kitchenham, Budgen & Brereton, 2016.  
Capítulo 7

# Objetivos de aprendizaje

---

## **Evaluación de calidad de los estudios**

---

34. Analizar la necesidad de evaluar la calidad de los estudios primarios

---

35. Explicar los conceptos y actividades asociados a la evaluación de calidad de estudios primarios

---

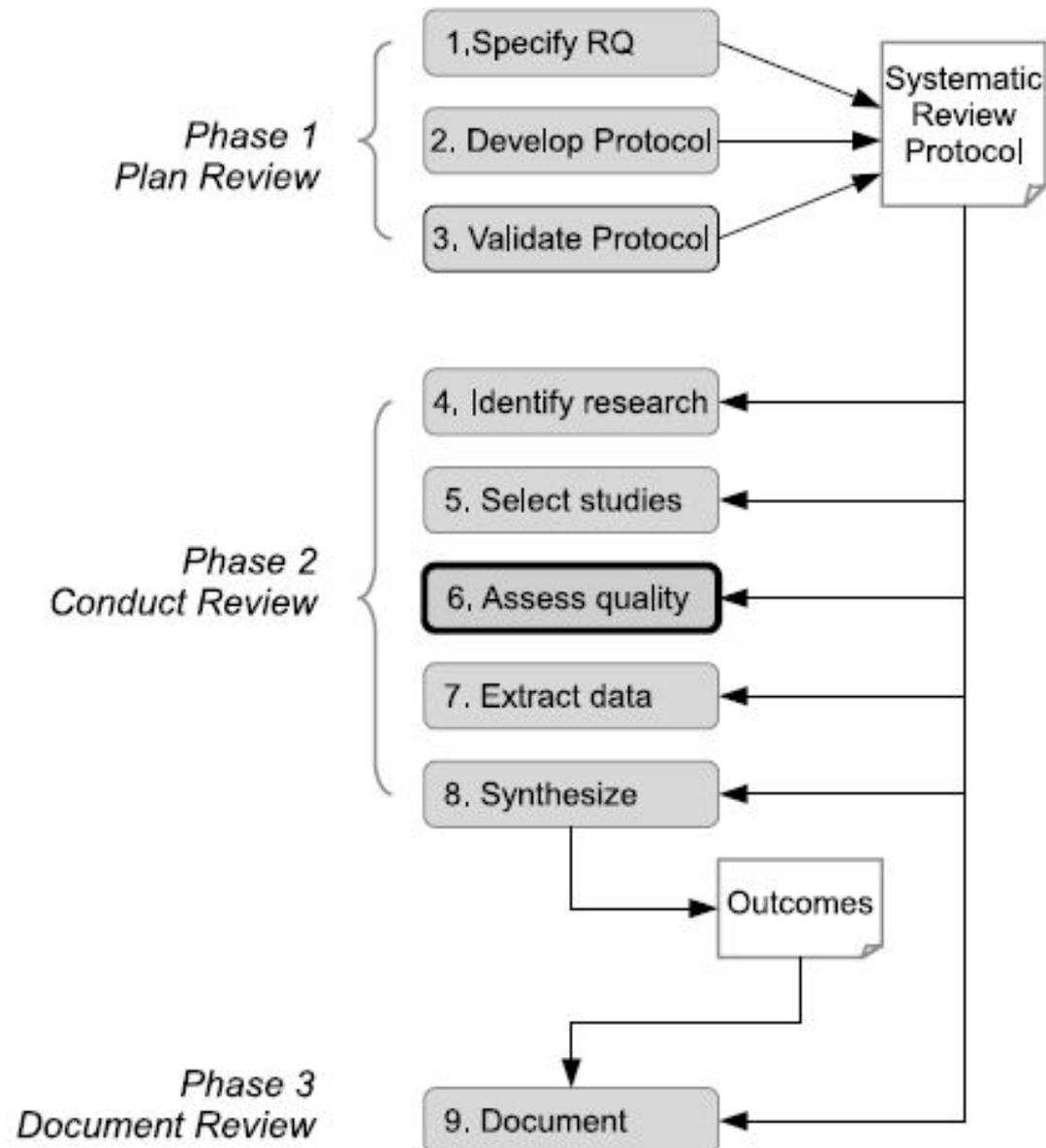
36. Participar en la definición de criterios de evaluación de calidad de estudios primarios para una SLR

---

37. Participar en la evaluación de calidad de estudios primarios de una SLR

---

# Proceso de revisión sistemática



# Evaluación de calidad

---

## → **¿Cuál es la necesidad de evaluar la calidad?**

- ◆ Evaluar la calidad con la que la evidencia fue obtenida en los estudios primarios
    - Guiar la interpretación de los hallazgos y la fortaleza de las inferencias
  - ◆ Proveer criterios más detallados de inclusión y exclusión
  - ◆ Investigar cuándo la diferencia de calidad explica diferencias en los resultados de los estudios
  - ◆ Poder ponderar el impacto de cada estudio al momento de sintetizar los resultados de la revisión
  - ◆ Guiar recomendaciones a futuras investigaciones
-

# Evaluación de calidad

---

## → Criterio de evaluación de calidad

- ◆ Los investigadores deben construir *checklists* de evaluación de calidad para evaluar cada estudio
  - El propósito de la evaluación de calidad es el que guía la construcción de las *checklists*
  - Se suelen construir escalas numéricas
- ◆ Hay un acuerdo sugerido acerca de que la calidad está relacionada con la forma en que los estudios minimizan los sesgos y maximizan la validez interna y externa



# Evaluación de calidad

---

- Conceptos

<b>Término</b>	<b>Sinónimo</b>	<b>Definición</b>
Sesgo	Error sistemático	Una tendencia a producir resultados alterados de forma sistemática a partir de los resultados verdaderos.
Validez interna	Validez	La medida en que el diseño y la ejecución del estudio previenen errores sistemáticos. La validez interna es un prerrequisito de la validez externa
Validez externa	Generalización, aplicabilidad.	La medida en que los efectos observados en el estudio son aplicables fuera del estudio

---

# Evaluación de calidad

---

## → Procedimiento de evaluación de calidad

- ◆ En un mapeo sistemático la evaluación de calidad no es obligatoria
  - ◆ La evaluación de calidad se lleva a cabo según los instrumentos definidos en el protocolo
  - ◆ Puede ser necesario excluir un estudio primario por no alcanzar el nivel mínimo de calidad requerido
  - ◆ La evaluación de calidad se realiza luego de aplicar el proceso de selección de estudios y antes de extraer la información
    - En general, estas actividades se retroalimentan
    - Puede realizarse al mismo momento en que se extrae información
-

# Evaluación de calidad

---

## Procedimiento de evaluación de calidad

### Puntaje

Puntuar según la *checklist* adecuada al tipo de estudio

### Validación de puntajes

Proceso de resolución a partir de puntajes de varios revisores

Si es un revisor solo, puede volver a realizar la evaluación en otro día (test-retest)

### Uso de los resultados de la evaluación de calidad

Excluir estudios de baja calidad

Analizar incluyendo los estudios de baja calidad y sin ellos para determinar el impacto en los resultados generales

Una pregunta de investigación puede abordar la calidad de los estudios encontrados

---

# Evaluación de calidad

---

## Proceso de resolución a partir de puntajes de varios revisores

### Moderación

Una tercera persona realiza la evaluación del estudio y se discute con los revisores en desacuerdo

Se propone que los revisores en desacuerdo trabajen juntos para llegar a un acuerdo

Combinar los puntajes, obtener la media

### Plan de contingencia

Convocar una reunión del equipo de revisores para discutir los criterios y las razones de los desacuerdos

Entrenamiento adicional para algún integrante del equipo

---

# SR Ejemplo

## 4. Proceso de Evaluación de la Calidad de los Estudios

<b>4.1. Se evaluará la calidad de los estudios (justificar)</b>	Si
<b>4.2. Checklist propuesta</b>	Ver en hoja "Evaluación de calidad"
<b>4.3. Cómo se evaluará el acuerdo entre revisores</b>	Lo harán los dos revisores. Algunas preguntas pueden aplicar y otras no, para estas últimas decir explícitamente "N/A".
<b>4.4. Cómo se resolverán diferencias</b>	Para las preguntas numéricas se hará promedio. Diferencias en los N/A serán analizadas utilizando correlación de Pearson y discutidas, en caso de dudas se pondrá N/A. Como la pregunta 2 es medio crítica, se tratará de usar algún mecanismo estadístico similar a Kappa para ver el nivel de acuerdo.
<b>4.5. Cómo se usarán las checklists</b>	Ver en hoja "Evaluación de calidad"

# SR Ejemplo - Protocolo

## Evaluación de calidad

Questions 3 through 12 admit the following answers: Yes / Partly / No / Not applicable. Score as 1, 0.5, 0. Interpolation is permitted for numerical values).

## Preguntas

1. Is the paper based on research (or is it a discussion paper based on expert opinion)? Yes / No.	
2. What research method was used: Experiment, Quasi-Experiment, Lessons learned, Case study, Opinion Survey, Other (specify)?	
Note: This is to be based on paper reading, not the method claimed by the authors.	
3. Is there a clear statement of the aims of the study?	
4. Is there an adequate description of the context in which the research or observation was carried out?	
5. Was the research method appropriate to address the aims of the research?	
6. Was the recruitment strategy (for human-based experiments and quasi-experiments) or experimental material or context (for Lessons learned) appropriate to the aims of the research?	
7. For empirical studies (apart from Lessons learned), was there a control group or baseline with which to evaluate SLR procedures/techniques?	
8. For empirical studies (apart from Lessons learned), was the data collected in a way that addressed the research issue?	
9. For empirical studies (apart from Lessons learned), was the data analysis sufficiently rigorous?	
10. Has the relationship between researcher and participants been considered to an adequate degree?	
11. Is there a clear statement of findings?	
12. Is the study of value for research or practice?	

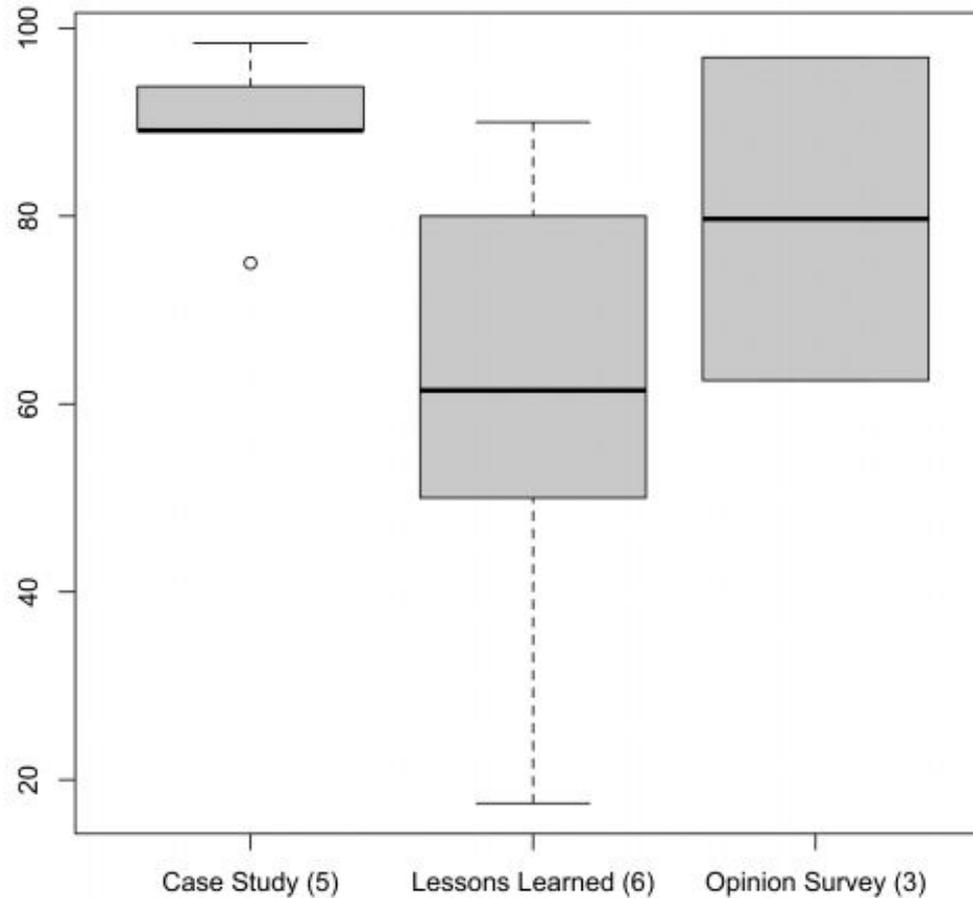
Study	1. Is the paper based on research (or is it a discussion paper based on expert opinion)? Yes/No	2. What research method was used: Experiment, Quasi-Experiment, Lessons learnt, Case study, Opinion Survey, Tertiary Study, Other (specify)? Note This is to be based on our reading of the paper not the method claimed by the author of the paper.  <i>if a case study was based on an opinion survey we called it "Opinion survey" rather than "Case Study", and if a study was based on a post-hoc re-analysis of a SR we called it an "Example" not a "Case Study" keeping the term "Case study" for an evaluation that was performed as part of undertaking an SR. We also identified very small experiments (e.g. 4 or fewer subjects) and small examples (e.g. one that only considered a part of the relevant data set or a small part of a specific task).</i>	3. Is there a clear statement of the aims of the study? Yes/Partly/No. Score as 1, 0,5, 0.	4. Is there an adequate description of the context in which the research or observation was carried out? Yes/Partly/No. Score as 1, 0,5, 0.
Seba	5	Opinion Survey	1	0,75
Sil	5	Case study	1	1
			1	0,875

# SR Ejemplo - Conducción

#	Study type	Overall quality	
1	Case study	89,1	$100 \times 7,1/8 = 89,1$
2	Opinion Survey	62,5	$100 \times 5/8 = 62,5$
3	Lessons learnt	80	$100 \times 4/5 = 80$
4	Case study	98,4	$100 \times 7,9/8 = 98,4$
5	Opinion Survey	96,9	$100 \times 7,8/8 = 96,9$
6	Opinion Survey	65,6	$100 \times 5,3/8 = 65,6$
7	Case study	89,1	$100 \times 7,1/8 = 89,1$
8	Lessons learnt	70,8	$100 \times 4,3/6 = 70,8$
9	Case study	75	$100 \times 6/8 = 75$
10	Lessons learnt	50	$100 \times 3/6 = 50$
11	Lessons learnt	17,5	$100 \times 0,9/5 = 17,5$
12	Lessons learnt	52,1	$100 \times 3,1/6 = 52,1$
13	Case study	93,8	$100 \times 7,5/8 = 93,8$
14	Lessons learnt	90	$100 \times 4,5/5 = 90$

# SR Ejemplo - Reporte

---



# Puntos clave

---

- El aspecto subjetivo de muchos criterios de calidad hace que el proceso de evaluación de calidad diste de ser simple
  - Se debe ser muy claro acerca de cómo será utilizada la evaluación de calidad
  - Si se tienen estudios primarios de distintos tipos y se utiliza una *checklist* genérica, además de la evaluación de calidad se debe tomar nota del tipo de estudio para analizar la confiabilidad de los resultados individuales.
  - Se deberían publicar estadísticas de acuerdos de forma de señalar la confiabilidad del proceso de evaluación de calidad
-

# Otro ejemplo de reporte

---

**Table 1: Reported benefits of academic EBSE trainings**

Benefit	Reported as results in
Acquire or improve research skills	[17, 33]
Become aware of the value of aggregating evidence	[5]
Learn how to search the literature and organize results	[33]
<i>Learn how to assess the relevance, validity or quality of the information on a topic</i>	[17]
<i>Practice the use of digital libraries</i>	[17]

Catal [17] did not indicate how he obtained these results.