



## CARTA EDITORIAL DE CORRECCIÓN



**La clase grabada como dispositivo evaluativo:  
Meta-evaluación de las rúbricas del Módulo 2 de la  
Evaluación Docente chilena**

The Recorded Lesson as an Evaluative Device:  
A Meta-Evaluation of the Rubrics of Module 2 in the Chilean  
Teacher Evaluation System

Auristela del Carmen Hormazabal Soto

Universidad de Playa Ancha (Chile)

Correo electrónico: [auristela.hormazabal@upla.cl](mailto:auristela.hormazabal@upla.cl)

ORCID: <https://orcid.org/0000-0002-8526-5109>

1

Mario Suárez Cinto

Universidad de Playa Ancha (Chile)

Correo electrónico: [mario.suarez@alumnos.upla.cl](mailto:mario.suarez@alumnos.upla.cl)

ORCID: <https://orcid.org/0009-0002-9423-2752>

**Resumen**

El presente estudio tiene como objetivo analizar las rúbricas del Módulo 2 (clase grabada) de la Evaluación del Desempeño Profesional Docente chilena como instrumento evaluativo, examinando su coherencia interna, su grado de observabilidad y su potencial formativo, con el fin de comprender la concepción de enseñanza y de “buena clase” que estas promueven. La investigación se desarrolló



## ARTÍCULO

desde un enfoque cualitativo, mediante un diseño de análisis documental evaluativo orientado a la meta-evaluación de instrumentos. No se consideraron participantes humanos; la unidad empírica estuvo constituida por las rúbricas oficiales del Módulo 2 correspondientes a los niveles de 7° básico a 4° medio, incluyendo indicadores, descriptores de desempeño y notas aclaratorias.

El análisis se realizó a partir de una matriz evaluativa construida sobre tres criterios: coherencia interna, observabilidad de los descriptores y potencial formativo. Los resultados muestran que las rúbricas presentan una coherencia estructural general, pero evidencian diferencias relevantes en la construcción de las progresiones evaluativas entre niveles de desempeño. En varios indicadores, dichas progresiones se articulan principalmente a partir de variaciones cuantitativas, lo que limita la explicitación de transformaciones cualitativas en la práctica pedagógica. Asimismo, se identifican niveles heterogéneos de observabilidad y un potencial formativo desigual entre indicadores.

2

En conclusión, el estudio evidencia tensiones entre la orientación formativa declarada por la política de evaluación docente y el diseño evaluativo concreto de las rúbricas del Módulo 2, aportando insumos para la revisión y fortalecimiento de instrumentos de evaluación docente orientados al desarrollo profesional.

**Palabras Clave:** evaluación docente; rúbricas; evaluación formativa; meta-evaluación.



### Abstract

This study examines the rubrics used in Module 2 (recorded lesson) of the Chilean Teacher Performance Evaluation System from an evaluative perspective. The aim is to analyze their internal coherence, degree of observability, and formative potential in order to understand the underlying conception of teaching and “good classroom practice” that they promote. The research adopts a qualitative approach, using an evaluative documentary analysis design oriented toward the meta-evaluation of assessment instruments. No human participants were involved. The empirical corpus consisted of the official Module 2 rubrics applied to grades 7 through 12, including performance indicators, level descriptors, and accompanying explanatory notes.

Data were analyzed through an evaluative matrix built around three analytical criteria: internal coherence, observability of descriptors, and formative potential. The findings indicate that, while the rubrics show a generally consistent structural organization, there are notable differences in how performance progressions are constructed across indicators. In several cases, progressions rely primarily on quantitative or frequency-based distinctions, which limits the explicit articulation of qualitative changes in teaching practice. Additionally, the results reveal heterogeneous levels of observability, with some descriptors requiring substantial inferential judgment, as well as uneven formative potential across indicators.



Overall, the study highlights tensions between the formative orientation declared by Chilean teacher evaluation policy and the concrete evaluative design of the Module 2 rubrics. These findings provide relevant insights for revising teacher evaluation instruments and strengthening their capacity to support professional learning.

**Keywords:** teacher evaluation; assessment rubrics; formative assessment; meta-evaluation.

### 1. Introducción

La evaluación del desempeño docente se ha consolidado como un componente central de las políticas educativas orientadas al aseguramiento de la calidad de la enseñanza y al fortalecimiento de la profesión docente. Desde la teoría de la evaluación, estos sistemas no solo cumplen una función certificadora, sino que operan como dispositivos reguladores que definen estándares de calidad, modelan prácticas pedagógicas y orientan procesos de desarrollo profesional (Scriven, 1991; Stufflebeam & Shinkfield, 2007). En este marco, los instrumentos evaluativos adquieren un rol estratégico, en tanto median entre la práctica observada y el juicio evaluativo emitido, configurando marcos normativos sobre lo que se considera una enseñanza de calidad (Scriven, 1991; Sadler, 1989; Brookhart, 2013).

En Chile, el Sistema de Evaluación del Desempeño Profesional Docente se inscribe en esta lógica, declarando explícitamente una orientación formativa centrada en la mejora de la práctica pedagógica y en el aprendizaje profesional del



## ARTÍCULO

profesorado (Ministerio de Educación de Chile [MINEDUC], 2023). Dentro de este sistema, el Portafolio de desempeño pedagógico constituye un instrumento clave, particularmente a través del Módulo 2, que evalúa la enseñanza mediante la observación de una clase grabada y el uso de rúbricas oficiales definidas por el Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (CPEIP). Estas rúbricas se alinean normativamente con el Marco para la Buena Enseñanza y buscan operacionalizar criterios de calidad pedagógica observables en la práctica docente (CPEIP, 2025)

No obstante, desde una perspectiva evaluativa, la evaluación de la práctica pedagógica a partir de una única situación de enseñanza plantea desafíos técnicos relevantes, especialmente cuando los criterios evaluativos remiten a constructos complejos, relacionales o ético-pedagógicos cuya observabilidad directa es limitada (Scriven, 1991; Darling-Hammond, 2013; Kane & Staiger, 2012). Diversos estudios advierten que la observación de clases como estrategia evaluativa implica necesariamente procesos de inferencia por parte del evaluador, lo que incrementa la carga interpretativa y tensiona la validez del juicio evaluativo cuando no existe una alineación clara entre criterios, evidencias y procedimientos de evaluación (Scriven, 1991; Stufflebeam & Shinkfield, 2007; Darling-Hammond, 2013).

A pesar de la centralidad del Módulo 2 en el sistema de evaluación docente chileno, la literatura disponible se concentra principalmente en la descripción normativa y operativa del instrumento, existiendo escasa evidencia empírica publicada que examine críticamente las rúbricas utilizadas desde una perspectiva evaluativa, interrogando su coherencia interna, su grado de observabilidad y su



## ARTÍCULO

potencial formativo (MINEDUC, 2023; CPEIP, 2025; Scriven, 1991). Asimismo, la documentación oficial se concentra principalmente en la descripción normativa de los criterios e indicadores, entregando información acotada respecto de cómo los descriptores permiten distinguir progresivamente niveles de calidad pedagógica y orientar procesos de mejora de la práctica (MINEDUC, 2023; CPEIP, 2025). Desde la teoría de la evaluación, esta ausencia de análisis sistemático del instrumento constituye un vacío relevante, en tanto toda evaluación y sus dispositivos deben ser, a su vez, objeto de evaluación para asegurar su calidad, utilidad y legitimidad (Scriven, 1991; Stufflebeam & Shinkfield, 2007).

Frente a este escenario, el presente estudio propone un análisis meta-evaluativo de las rúbricas del Módulo 2 de la Evaluación Docente chilena, con el propósito de aportar evidencia teórica y técnica para la comprensión crítica del diseño evaluativo del instrumento y de la concepción de enseñanza que este promueve, contribuyendo al debate sobre evaluación docente en contextos de alto impacto y a la mejora futura de instrumentos orientados al desarrollo profesional docente (Scriven, 1991; Stufflebeam & Shinkfield, 2007).

6

### 1.1. Pregunta de investigación

¿En qué medida las rúbricas del Módulo 2 de la Evaluación Docente chilena presentan coherencia interna, observabilidad y potencial formativo, y qué concepción de enseñanza y de “buena clase” se configura a partir de sus criterios y descriptores evaluativos?



### 1.2. Objetivo general

Analizar las rúbricas del Módulo 2 (clase grabada) de la Evaluación Docente chilena como instrumento evaluativo, examinando su coherencia interna, su grado de observabilidad y su potencial formativo, con el fin de comprender la concepción de enseñanza y de práctica pedagógica que estas promueven.

### 1.3. Objetivos específicos

- i. Analizar la coherencia interna de las rúbricas del Módulo 2, considerando la consistencia entre criterios, descriptores y niveles de desempeño.
- ii. Examinar el grado de observabilidad de los descriptores evaluativos en relación con la evidencia disponible a partir de una clase grabada.
- iii. Comprender el potencial formativo de las rúbricas del Módulo 2 y la concepción de enseñanza y de “buena clase” que estas configuran, a partir del análisis de sus progresiones evaluativas y descriptores de desempeño.

## 2. Metodología

### 2.1. Método

La investigación se desarrolló desde un enfoque cualitativo, mediante un diseño de análisis documental evaluativo, orientado a la meta-evaluación de un instrumento de evaluación docente. La meta-evaluación se comprende como el proceso



sistemático de examinar la calidad, coherencia y utilidad de una evaluación o de sus instrumentos, con el propósito de emitir juicios fundamentados sobre su valor evaluativo y formativo (Scriven, 1991; Stufflebeam & Shinkfield, 2007).

Este método resulta pertinente en tanto el estudio no busca evaluar el desempeño de docentes individuales, sino analizar críticamente el diseño, la estructura y los supuestos evaluativos que subyacen a las rúbricas del Módulo 2 (clase grabada) de la Evaluación Docente chilena, entendidas como dispositivos normativos que regulan y orientan la práctica pedagógica.

### 2.2. Participantes

Dado el carácter cualitativo y documental del estudio, no se consideraron participantes humanos. La unidad empírica de análisis estuvo constituida por documentos normativos oficiales del Sistema de Evaluación del Desempeño Profesional Docente en Chile. Específicamente, se analizaron las rúbricas del Módulo 2 (clase grabada), correspondientes a los niveles de 7° básico a 4° medio.

El corpus documental incluyó los indicadores evaluativos del módulo, los descriptores de desempeño asociados a los niveles Insatisfactorio, Básico, Competente y Destacado, las notas aclaratorias y orientaciones técnicas para el proceso de corrección, así como las referencias a las fuentes de evidencia consideradas en la evaluación (grabación de la clase, ficha de la clase grabada y recursos adjuntos).



### 2.3. Estrategias

La estrategia de investigación utilizada fue el análisis documental evaluativo. Para ello, se diseñó una matriz evaluativa como instrumento de análisis, que permitió examinar sistemáticamente cada descriptor de desempeño en función de criterios evaluativos previamente definidos. Esta estrategia posibilitó un análisis estructurado y comparativo de los descriptores, resguardando la coherencia entre los objetivos del estudio y el procedimiento analítico.

### 2.4. Procedimiento

El procedimiento de análisis se desarrolló en varias etapas. En primer lugar, se realizó una lectura analítica exhaustiva del corpus documental, con el fin de identificar la estructura general de las rúbricas y sus componentes evaluativos. En segundo lugar, se efectuó un análisis descriptor por descriptor, registrando observaciones analíticas y asignando categorías evaluativas justificadas según los criterios definidos. En una tercera etapa, se elaboraron notas analíticas por indicador, integrando los hallazgos parciales a nivel interpretativo. Finalmente, se construyó una síntesis transversal de los resultados.

Con el propósito de resguardar el rigor científico del estudio, se consideraron criterios de credibilidad, transparencia y trazabilidad del análisis, asegurando la coherencia entre los fundamentos teóricos, el procedimiento seguido y los resultados obtenidos (Lincoln & Guba, 1985). Dado que se trabajó exclusivamente



con documentos públicos y normativos, no fue necesario aplicar procedimientos de resguardo ético asociados a participantes humanos.

El análisis detallado de cada descriptor se realizó mediante una matriz evaluativa diseñada específicamente para este estudio, la cual permitió registrar de manera sistemática los juicios evaluativos asociados a los criterios de coherencia interna, observabilidad y potencial formativo. Considerando el carácter exhaustivo de dicha matriz y las limitaciones de extensión propias del formato artículo, se optó por incluir en los anexos únicamente un ejemplo representativo del análisis de un indicador. La matriz evaluativa completa fue utilizada como respaldo metodológico y evidencia de la trazabilidad del proceso analítico, encontrándose disponible para su revisión previa solicitud.

### 2.5. Análisis de la información

El análisis de la información se realizó mediante un análisis documental evaluativo, con apoyo de procedimientos propios del análisis de contenido cualitativo. Cada descriptor fue analizado como una unidad textual evaluativa, atendiendo a los objetivos del estudio y a los criterios de análisis definidos.

Los criterios utilizados fueron: coherencia interna, entendida como la correspondencia lógica y pedagógica entre indicador, descriptor y nivel de desempeño; observabilidad, referida a la posibilidad de evidenciar empíricamente los descriptores a partir de una clase grabada; y potencial formativo, entendido



como la capacidad del instrumento para orientar la reflexión pedagógica y la mejora de la práctica docente (Sadler, 1989; Brookhart, 2013; Stufflebeam & Shinkfield, 2007).

### 3. Resultados

El análisis meta-evaluativo de las rúbricas del Módulo 2 permitió identificar patrones consistentes en relación con la coherencia interna del instrumento, el grado de observabilidad de los descriptores y su potencial formativo. Los resultados se presentan organizados según estos tres criterios analíticos, considerando el conjunto de indicadores que componen la rúbrica.

11

#### 3.1. Coherencia interna de las rúbricas del Módulo 2

En términos de coherencia interna, el análisis evidenció que la mayoría de los indicadores presenta una estructura formal consistente entre criterio, descriptor y nivel de desempeño. No obstante, se observaron diferencias relevantes en la forma en que se construyen las progresiones evaluativas entre los niveles Insatisfactorio, Básico, Competente y Destacado.

En varios indicadores, la progresión entre niveles se articula principalmente a partir de variaciones de frecuencia, extensión temporal o cobertura de la acción pedagógica, sin explicitar transformaciones cualitativas claras en la práctica docente



evaluada. En estos casos, los descriptores mantienen una estructura similar entre niveles, diferenciándose fundamentalmente por expresiones tales como “en algunos momentos”, “durante la mayor parte de la clase” o “de manera sistemática”.

En contraste, algunos indicadores, particularmente aquellos asociados a dimensiones como equidad de género, incorporan descriptores que explicitan acciones pedagógicas diferenciadas entre niveles, lo que permite identificar con mayor claridad una progresión cualitativa del desempeño evaluado.

### 3.2. Observabilidad de los descriptores evaluativos

Respecto del grado de observabilidad, los resultados muestran que los descriptores presentan distintos niveles de accesibilidad empírica a partir de una clase grabada. Se identificaron descriptores con observabilidad directa, vinculados a acciones pedagógicas visibles en la interacción de aula, tales como la organización de actividades, la promoción de la participación estudiantil o el uso de recursos didácticos.

Sin embargo, una proporción significativa de los descriptores remite a constructos relacionales, disposicionales o intencionales, cuya evidencia requiere inferencia por parte del evaluador. En estos casos, la observabilidad se clasificó como inferencial, dado que la verificación del descriptor depende de la interpretación de gestos, interacciones o decisiones pedagógicas que no siempre se manifiestan explícitamente en una única situación de enseñanza.



Asimismo, se identificaron descriptores cuya observabilidad resulta limitada o no observable directamente en una clase grabada, especialmente aquellos asociados a procesos de desarrollo progresivo, disposiciones ético-pedagógicas o efectos de largo plazo en el aprendizaje de los estudiantes.

### 3.3. Potencial formativo del instrumento

En relación con el potencial formativo, los resultados indican que las rúbricas del Módulo 2 presentan un nivel variable de orientación a la mejora de la práctica docente. En general, los descriptores permiten identificar el nivel de desempeño alcanzado, pero ofrecen información acotada respecto de las acciones pedagógicas específicas necesarias para avanzar hacia niveles superiores.

13

Las notas aclaratorias asociadas a los indicadores contribuyen a ampliar la comprensión conceptual de los criterios evaluados. No obstante, estas orientaciones no siempre se integran de manera explícita en los descriptores de desempeño, lo que limita su función orientadora en términos formativos.

De manera particular, los indicadores que incorporan descriptores con acciones pedagógicas explícitas y diferenciadas entre niveles evidencian un mayor potencial para orientar la reflexión docente y la toma de decisiones didácticas, en comparación con aquellos basados predominantemente en variaciones cuantitativas.



### 3.4. Síntesis de resultados

Con el fin de sintetizar los principales hallazgos del análisis meta-evaluativo, la Tabla 1 presenta un resumen de los resultados obtenidos para cada indicador de la rúbrica, considerando los criterios de coherencia interna, observabilidad y potencial formativo.

**Tabla 1:** Síntesis del análisis evaluativo de las rúbricas del Módulo 2

Indicador	Progresión evaluativa predominante	Observabilidad	Potencial formativo
Ambiente de aula	Media	Inferencial	Medio
Promoción de la participación	Media	Mixta	Medio
Contribución de las actividades al logro de los objetivos	Media	Directa	Medio
Promoción de habilidades para el aprendizaje profundo	Baja-Media	Inferencial	Bajo
Uso de conocimientos y experiencias de los/as estudiantes	Media	Inferencial	Medio
Uso de los desempeños para favorecer el aprendizaje	Media	Directa	Medio
Promoción de la equidad de género	Alta	Mixta	Alto



Nota. La clasificación presentada se construyó a partir del análisis sistemático de los descriptores y notas aclaratorias de cada indicador, considerando el conjunto del instrumento y no juicios aislados por descriptor

En conjunto, los resultados muestran que las rúbricas del Módulo 2 presentan variaciones entre indicadores en términos de coherencia interna, observabilidad de los descriptores y potencial formativo, según los criterios de análisis considerados.

#### 4. Discusión

Los resultados del análisis meta-evaluativo de las rúbricas del Módulo 2 permiten discutir el diseño del instrumento desde una perspectiva evaluativa, atendiendo a su coherencia interna, su grado de observabilidad y su potencial formativo. A continuación, se analizan los hallazgos en relación con los objetivos y preguntas de investigación, se comparan con la literatura especializada, se interpretan a la luz de marcos teóricos de la evaluación educativa y se examinan sus implicancias para el diseño de instrumentos de evaluación docente.

##### 4.1. Relación de los resultados con los objetivos y preguntas de investigación

El objetivo general del estudio fue analizar las rúbricas del Módulo 2 de la Evaluación Docente chilena como instrumento evaluativo, examinando su coherencia interna, observabilidad y potencial formativo, con el fin de comprender



la concepción de enseñanza y de “buena clase” que estas promueven. En este sentido, los resultados muestran que las rúbricas presentan niveles diferenciados en los tres criterios analizados, lo que permite responder afirmativamente a la pregunta de investigación planteada, evidenciando que el diseño evaluativo del instrumento no es homogéneo entre indicadores.

En particular, el análisis reveló que, si bien existe una coherencia estructural básica entre criterios, descriptores y niveles de desempeño, en varios indicadores las progresiones evaluativas se construyen principalmente a partir de variaciones cuantitativas o temporales, lo que limita la explicitación de transformaciones cualitativas en la práctica pedagógica. Asimismo, los resultados muestran que el grado de observabilidad de los descriptores es heterogéneo, oscilando entre acciones directamente observables y constructos que requieren inferencia interpretativa. Finalmente, el potencial formativo del instrumento se manifestó de manera desigual, siendo mayor en aquellos indicadores que explicitan acciones pedagógicas diferenciadas entre niveles de desempeño.

#### 4.2. Comparación con estudios y marcos teóricos descritos en la introducción

Estos hallazgos son consistentes con la literatura especializada sobre evaluación del desempeño docente y diseño de rúbricas. Diversos autores han advertido que las rúbricas que basan sus progresiones principalmente en criterios de frecuencia o extensión tienden a reforzar una lógica evaluativa de carácter sumativo, incluso



## ARTÍCULO

cuando se presentan discursivamente como instrumentos formativos (Sadler, 1989; Brookhart, 2013). En este sentido, los resultados del estudio coinciden con lo señalado por Brookhart (2013), quien sostiene que las rúbricas con progresiones cuantitativas ofrecen información limitada para orientar procesos de mejora de la práctica.

Asimismo, la heterogeneidad observada en el grado de observabilidad de los descriptores se alinea con los planteamientos de Scriven (1991) y Darling-Hammond (2013), quienes advierten que la observación de clases como estrategia evaluativa implica necesariamente procesos de inferencia cuando los criterios evaluados remiten a constructos relacionales o disposicionales. En el caso del Módulo 2, esta situación resulta especialmente relevante dado que varios indicadores se orientan a dimensiones como clima de aula, relación pedagógica o equidad de género, las cuales no siempre se manifiestan de manera directa en una única situación de enseñanza.

Por otra parte, los resultados dialogan con los planteamientos de Stufflebeam y Shinkfield (2007) respecto de la necesidad de asegurar una alineación clara entre criterios, evidencias y procedimientos de evaluación para resguardar la validez del juicio evaluativo. En aquellos indicadores donde esta alineación se presenta de forma más explícita, el instrumento muestra mayores niveles de coherencia interna y potencial formativo.



### 4.3. Interpretación de los resultados

Desde una perspectiva evaluativa, los resultados sugieren que las rúbricas del Módulo 2 configuran un instrumento que combina elementos propios de una racionalidad certificadora con componentes asociados a una orientación formativa. La presencia de progresiones cuantitativas predominantes en varios indicadores indica que el juicio evaluativo se construye, en gran medida, a partir de la constatación de la ocurrencia o extensión de determinadas prácticas, más que de la explicitación de cómo estas prácticas pueden transformarse pedagógicamente.

Al mismo tiempo, la incorporación de notas aclaratorias y descriptores más cualitativos en algunos indicadores introduce elementos que amplían la comprensión del criterio evaluado, aunque estos no siempre se integran de manera explícita en la progresión de niveles. Esta tensión entre definición conceptual y operacionalización evaluativa refuerza la idea de que el diseño de las rúbricas cumple una función reguladora del desempeño docente, comunicando estándares de calidad, pero con un alcance formativo limitado en ciertos ámbitos.

La variabilidad observada entre indicadores sugiere, además, que el instrumento no responde a una única concepción de enseñanza, sino que articula distintas visiones sobre lo que constituye una “buena clase”, combinando énfasis en la gestión del aula, la participación estudiantil, el desarrollo de habilidades cognitivas y la dimensión ética de la enseñanza.



#### 4.4. Implicancias del estudio

Las implicancias del estudio son relevantes tanto para la revisión del Módulo 2 como para el diseño futuro de instrumentos de evaluación docente. En primer lugar, los resultados sugieren la necesidad de fortalecer progresiones evaluativas cualitativas más explícitas en los niveles intermedios de desempeño, de modo que las rúbricas orienten con mayor claridad los procesos de mejora de la práctica pedagógica.

En segundo lugar, se evidencia la conveniencia de integrar de manera más directa las notas aclaratorias en los descriptores evaluativos, reduciendo la distancia entre la definición conceptual del criterio y el juicio evaluativo que se emite a partir de la evidencia observada. Asimismo, los hallazgos destacan la importancia de revisar la alineación entre criterios y fuentes de evidencia, con el fin de disminuir la carga inferencial del evaluador, especialmente en indicadores asociados a constructos complejos.

Finalmente, el estudio aporta evidencia para comprender el Módulo 2 como un dispositivo evaluativo con efectos normativos sobre la enseñanza, lo que refuerza la necesidad de someter estos instrumentos a procesos sistemáticos de meta-evaluación. En este sentido, los resultados contribuyen al debate sobre la evaluación docente en contextos de alto impacto y ofrecen insumos técnicos para el fortalecimiento de políticas de evaluación orientadas al desarrollo profesional docente.



### Bibliografía

Black, P., & Wiliam, D. (2009). *Developing the theory of formative assessment*.  
*Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.

<https://doi.org/10.1007/s11092-008-9068-5>

Brookhart, S. M. (2013). How to create and use rubrics for formative assessment and grading. ASCD.

Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (CPEIP). (2025). Rúbricas del Módulo 2: Clase grabada. Evaluación del desempeño profesional docente. Ministerio de Educación de Chile.

Darling-Hammond, L. (2013). Getting teacher evaluation right: What really matters for effectiveness and improvement. Teachers College Press.

20

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Bill & Melinda Gates Foundation.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Ministerio de Educación de Chile (MINEDUC). (2023). Sistema de evaluación del desempeño profesional docente. <https://www.docentemas.cl>

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.

<https://doi.org/10.1007/BF00117714>



## ARTÍCULO

Scriven, M. (1991). Evaluation thesaurus (4th ed.). Sage Publications.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). Evaluation theory, models, and applications. Jossey- Bass.