



Modelos de Evaluación de la Docencia en Educación Superior: experiencias y aplicaciones

Tamara Zuñiga, Enzo Zurita & Carlos González

ÍNDICE

Introducción	4
1. Criterios sobre Buena Docencia	5
1.1. Criterios para evaluar la docencia de calidad	6
a. Dimensión de estructura del curso	6
b. Dimensión de interacción social	7
c. Dimensión de habilidades de presentación	8
d. Dimensión de habilidades para la gestión docente	9
e. Dimensión de estrategias docentes	9
1.2. Criterios de evaluación que utilizan las Instituciones de Educación Superior	11
a. Dimensiones de logro en buena docencia definidas a través de marcos, modelos y/o criterios institucionales chilenos	11
b. Dimensiones de logro en buena docencia definidas a través de marcos, modelos y/o criterios institucionales internacionales	13
1.3. Comparación general de los criterios de buena docencia	16
1.4. Docencia en contextos virtuales	18
2. Instrumentos para la Evaluación de la Docencia	21
2.1. Cuestionario de Estudiantes Evaluando la Docencia (SET)	21
a. Definición	21
b. Aspectos positivos	21
c. Aspectos negativos	22
d. Tendencias emergentes	22
2.2. Observación de Aula	23
a. Definición	23
b. Aspectos positivos	25
c. Aspectos negativos	25
d. Tendencias emergentes	25
2.3. Portafolio docente	26
a. Definición	26
b. Aspectos positivos	27
c. Aspectos negativos	27

d. Tendencias emergentes	28
3. Ejemplos de Instrumentos de Evaluación Docente	29
a. Universidad de Melbourne	29
b. Universidad de Carolina del Norte	30
c. Universidad de Minnesota	30
d. Universidad de Stanford	31
e. Universidad de Oxford	31
f. Universidad de College London	32
4. Un ejemplo de convergencia entre los criterios e instrumentos de evaluación docente: El Marco de las Universidades Australianas	34
5. Alternativas para la implementación de los Instrumentos de Evaluación Docente	36
5.1. Características generales de los instrumentos de evaluación docente	36
5.2. Cuestionario de Estudiantes o SET.	37
5.3. Observación de Aula	40
5.4. Portafolio docente	42
6. Actores involucrados en el proceso de evaluación docente	47
6.1. Estudiantes	47
6.2. Pares y colegas evaluadores	48
6.3. Docentes en desarrollo	48
7. Reflexiones finales	50
Referencias	53

INTRODUCCIÓN

La evaluación docente, tal como mencionan Chalmers & Hunt (2016), es una práctica que busca la mejora de las oportunidades de aprendizaje de los estudiantes, de los planes de estudio y desarrollo de los mismos profesores en su carrera profesional. En particular, puede realizarse con fines formativos, tales como mejorar la enseñanza, pero también con fines sumativos, por ejemplo, determinar la competencia docente para la toma de decisiones administrativas al programar nuevos semestres (Spooren et al., 2013).

En la actualidad, evaluar a los docentes, se ha visto enmarcado en un contexto de exigencias desde los distintos gobiernos hacia las Instituciones de Educación Superior (IES) para cumplir con ciertos estándares de calidad, dónde a su vez se consideran otros elementos importantes tales como su plan de estudios, perfil de egreso del estudiante, organización y administración, etc. Como plantean Esterhazy et al. (2021), la calidad de la enseñanza es uno de los factores que más influyen en los resultados de aprendizaje y motivación en los estudiantes, de ahí que sea tan importante considerar estos distintos aspectos como fuente de evidencia.

Si bien, distintos instrumentos se han utilizado como base para la evaluación docente, como, por ejemplo, la autorreflexión, el cuestionario de estudiantes o revisión de pares. Desde hace un tiempo, diferentes autores han abogado por utilizar más de uno, para contar así con distintas fuentes para informar sobre la buena docencia, esto es, los profesores pares, el estudiante e inclusive el proceso mismo de enseñanza (Chalmers & Hunt, 2016; Felder & Brent, 2004). El argumento detrás de esto es que se podría obtener una imagen más completa de los diferentes aspectos y actividades que involucran la enseñanza en la Educación Superior (ES), que a su vez son fundamentales para evaluar de forma válida y fiable.

El siguiente documento tiene el objetivo de entregar información sobre los distintos aspectos que involucran la evaluación docente y que son relevantes para el diseño de un modelo integral de evaluación que considere los diferentes instrumentos y actores que están involucrados en el proceso de enseñanza. Esto también busca servir de guía para que las IES puedan producir mejoras y desarrollar buenas prácticas de enseñanza. Para ello, el documento cuenta con las siguientes secciones: 1. Criterios sobre Buena Docencia, 2. Instrumentos para la Evaluación de la Docencia, 3. Ejemplos de Instrumentos de Evaluación Docente, 4. Un ejemplo de convergencia entre los criterios e instrumentos de evaluación docente, 5. Alternativas de Implementación de los Instrumentos de Evaluación Docente, 6. Actores involucrados en el proceso de evaluación docente y 7. Reflexiones finales.

1. Criterios sobre Buena Docencia

En esta primera sección se indaga en los criterios para evaluar la buena docencia. Para ello, en primer lugar, se introduce con una discusión sobre cómo entender y evaluar a los profesores, y la relevancia de identificar las distintas dimensiones de su quehacer para tener una visión holística de su desempeño. En segundo lugar, se plantean 16 criterios para evaluar la calidad de la docencia, los que están agrupados en cinco grandes dimensiones, que a su vez están ordenadas según su relevancia e impacto en la enseñanza. En tercer lugar, se presentan competencias o criterios que se obtuvieron de la revisión de modelos de evaluación nacionales e internacionales, además, se ahonda en otros aspectos que estos presentan. Por último, en complemento a la primera y segunda sección, se muestra un apartado sobre docencia en contextos virtuales.

La evaluación de la enseñanza implica en primer lugar una discusión sobre qué entendemos por buena docencia y los distintos criterios que podemos emplear para identificarla y medirla en nuestras instituciones. Al respecto, Asún y Zúñiga (2017) indican que las encuestas que evalúan la docencia utilizan adjetivos como “bueno”, “competente”, “excelente” y/o “efectivo” para referirse a la buena enseñanza. En particular, los autores adscriben a la noción de “docente competente”, argumentando que el concepto de competencias permite identificar el adecuado ejercicio de la función docente en distintas, valga la redundancia, áreas de competencia.

En contraste a la perspectiva de Asún y Zúñiga (2017), está el concepto de “docencia de calidad”, al que nos adscribiremos en este informe. Esta noción, ha sido ampliamente utilizada en la literatura (Por ejemplo, Valencia, 2020; Spooren et al., 2013; Klemenčič & Chirikov, 2015) y nos permite observar los distintos criterios que indican la calidad y complejidad de la evaluación docente, más allá de las competencias que tienen los y las profesoras, entendiendo que hay diferentes dimensiones de su quehacer, permitiendo obtener una visión más holística de su desempeño.

Ahora bien, considerando la innovación tecnológica, asociada con la creciente accesibilidad del internet, la flexibilidad de las nuevas formas de transmitir conocimientos y contenidos (Kebritchi et al., 2016) y la llegada de la pandemia por COVID-19, la educación virtual se ha convertido en un aspecto integral de la educación superior. Por tanto, también se profundiza en el desafío de repensar la calidad de la docencia en diversos formatos.

Tratándose de un campo que ha obtenido creciente cobertura investigativa en las últimas décadas (Kebritchi et al., 2016), la docencia virtual puede ser asociada con diversos criterios de calidad, muchos de los cuales difieren de los elementos de la docencia presencial. Sin embargo, Pusey et al. (2021) demuestran que existen aspectos de ambas docencias que resultan lo suficientemente similares como para ser adaptados e incorporados integralmente en un modelo de evaluación docente que sea lo suficientemente pragmático y amplio, tal como se observa en su modelo E-COPUS.

Las siguientes secciones profundizan en los criterios que se han utilizado para evaluar la docencia de calidad. Además, en la segunda sección se ahonda en los criterios y/o competencias que utilizan distintas instituciones de educación superior alrededor del mundo.

1.1. Criterios para evaluar la docencia de calidad

En línea con la discusión sobre cómo entendemos la buena docencia, en esta sección se presentan un total de 16 criterios que permiten identificar y evaluar su calidad, los cuales están agrupados en cinco grandes dimensiones: a. Estructura del curso, b. Interacción social, c. Habilidades de presentación, d. Habilidades para la gestión docente y e. Estrategias de docentes.

Los criterios y dimensiones fueron ordenados en términos de su relevancia, de acuerdo con tres trabajos de investigación (Smith y Baik, 2019; Schneider y Preckel, 2017; Asún y Zúñiga, 2017). Estos son posteriormente analizados a la luz de investigaciones enfocadas en la docencia virtual (González et al., 2023; Orsini et al., 2022; Gómez-Rey et al., 2018).

a. Dimensión de estructura del curso

La **primera dimensión** en términos de relevancia agrupa la claridad en el diseño, desarrollo y socialización de las actividades del curso, así como el manejo docente del instructor a cargo de este. Bajo este marco, tres criterios pueden ser identificados:

- i. **Estructura:** Definida por Smith y Baik (2019) como la entrega al estudiantado de representaciones claras del contenido conceptual del curso y de la relación entre las ideas y conceptos que se están enseñando. De acuerdo con los autores, esto permite avanzar hacia un propósito bien definido y conocido por todos los actores involucrados en el proceso de enseñanza. Asún y Zúñiga (2017) consideran también este criterio bajo la denominación de “Planificación y Conducción del curso”, y le atribuyen funciones como la (1) presentación de los objetivos del programa, (2) la articulación adecuada y coherente de actividades y contenidos y (3) cumplir con la programación del curso, entre otros. Por su parte, Schneider y Preckel (2017) consideran este criterio de manera parcial, subsumiéndolo en la dimensión de “Evaluación” (pp. 588), específicamente en sus elementos relativos a la claridad de métodos y criterios de logro.
- ii. **Alineamiento:** Entendido por Smith y Baik (2019) como el diseño de actividades curriculares que apoyen el cumplimiento de logros y que se asimilen a las evaluaciones. Esta dimensión se encuentra presente también en el trabajo de Schneider y Preckel (2017) en su dimensión de “Evaluación”, donde sostienen que la mayor función de las evaluaciones es ofrecer a los estudiantes retroalimentación sobre sus progresos y las necesidades de su desarrollo futuro, poniendo énfasis en factores como; evaluación de pares, autoevaluación, claridad en los criterios evaluativos. Finalmente, Asún y Zúñiga (2017) asocian este criterio con la dimensión de “Evaluaciones válidas y pertinentes”, asignándole áreas de competencia como la producción de evaluaciones adecuadas, verificables y transparentes, congruentes con lo trabajado en el curso y susceptibles de utilizar como fuentes de retroalimentación para los estudiantes, entre otras.
- iii. **Expertiz:** Comprendido por Smith y Baik (2019) como el uso de docentes expertos por sobre docentes pares u otros profesionales no-expertos en sus áreas. Este elemento sólo se encuentra considerado en el trabajo de Asún y Zúñiga (2017), que asocian al “Guía Experto” con la capacidad de responder preguntas y sugerir bibliografía complementaria,

seleccionar y acentuar contenidos clave, exponer y discutir diversas perspectivas y aplicar contenidos a situaciones concretas, entre otros aspectos. Por otra parte, Schneider y Preckel (2017) no consideran la expertiz como parte de ninguna de las características atribuidas a una docencia de alto impacto.

En cuanto a la relevancia de los criterios, Smith y Baik (2019) identifican el elemento de la estructura como la práctica a la que se atribuye mayor impacto en la formación en educación superior. Además, observan que el alineamiento entre las actividades del curso y los métodos de evaluación corresponde a la segunda acción de mayor impacto en la formación. Finalmente, aunque se atribuye relevancia a la expertiz del docente que dicta el curso, este criterio adopta una posición intermedia en la escala de relevancia atribuida por la literatura. Por su parte, Schneider y Preckel (2017) indican que la estructura y alineamiento (bajo el marco de la variable “evaluación”) constituyen la tercera variable con mayor tamaño de efecto en la literatura revisada, por debajo de variables de carácter social, como interacción y estímulo del aprendizaje significativo. La confluencia de estas variables relativas al diseño y consistencia interna del curso sugiere que este primer criterio es uno de los más importantes para garantizar una docencia de calidad en educación superior.

b. Dimensión de interacción social

La **segunda dimensión** corresponde a aquellas prácticas que el docente puede desarrollar en consideración directa de su relación con el estudiantado de su curso.

- i. **Estimular el aprendizaje significativo:** La noción de “aprendizaje significativo” es definida por Asún y Zúñiga (2017) como la transmisión de valores, esquemas de interpretación y pautas de comportamiento propias de un profesional autónomo y eficaz. Es decir, este criterio considera al docente no sólo como un transmisor de contenidos curricularmente definidos, sino también un formador de profesionales a través del cual se reproducen las buenas prácticas propias de una disciplina específica. Schneider y Preckel (2017) consideran este criterio bajo el mismo nombre, indicando que para su correcta ejecución se requiere planificación y organización del curso, así como criterios y objetivos claros que puedan ser comprendidos y compartidos por el estudiantado. Además, los autores pueden aplicar esta práctica a través de la vinculación de los contenidos con experiencias de vida y metas del estudiantado, lo que les ayuda a percibir la relevancia del nuevo contenido, así como integrarlo con conocimientos anteriores.
- ii. **Motivación a los estudiantes:** Es definida como la capacidad de interesar a los estudiantes en los contenidos (Asún y Zúñiga, 2017), lo que se materializa a través de la capacidad de generar clases entretenidas, mostrar entusiasmo por los temas y la enseñanza, el planteamiento de problemas desafiantes y la explicación de la importancia de los contenidos. Por su parte, Schneider y Preckel (2017) abordan indirectamente este criterio, como parte de la dimensión de “interacción social”, asociada a estimular preguntas y discusiones, formular preguntas abiertas, como también la disponibilidad, amabilidad, preocupación y respeto por los estudiantes que el docente expresa.
- iii. **Interacción eficaz con estudiantes:** Finalmente, Asún y Zúñiga (2017) definen este criterio como la capacidad de escucha, respeto y adaptación a los estudiantes, asignándole

indicadores como la capacidad de diálogo, la detección de la diversidad y dificultades del estudiantado, la adaptación a sus demandas, expresión de interés por su aprendizaje y la mantención de la disciplina sin autoritarismo, entre otros aspectos. Schneider y Preckel (2017) consideran estos criterios en su dimensión de “interacción social” explicada en el punto anterior.

En términos de la relevancia de los criterios, de acuerdo con Schneider y Preckel (2017) las variables relativas a la interacción social y estímulo del aprendizaje significativo con estudiantes presentan los mayores tamaños de efecto en términos de impacto en la formación, afirmación respaldada por Asún y Zúñiga (2017), quienes identificaron un alto énfasis en estas habilidades por parte de la literatura de formación por competencias, los instrumentos previamente analizados y por el estudiantado. Por su parte, Smith y Baik (2019) le atribuyen el sexto lugar en términos de relevancia a la variable “Relación con el estudiantado”.

c. Dimensión de habilidades de presentación

Una **tercera dimensión** corresponde a aquellas prácticas o competencias que pueden ser desarrolladas por un docente al interior del aula para estimular el aprendizaje del estudiantado.

- i. **Riqueza episódica:** Definida por Smith y Baik (2019) como la entrega de ejemplos de aplicación de los conceptos tratados en el curso a través de anécdotas detalladas, específicas y reales sobre el desempeño profesional.
- ii. **Capacidad de comunicar contenidos:** Asún y Zúñiga (2017) definen esta competencia como la capacidad de expresarse con claridad en la transmisión de contenidos, asignándole indicadores como el desarrollo de esquemas explicativos, la graduación de la dificultad de los contenidos y reiteración de aquellos que resulten más difíciles, así como habilidades de expresión como voz, ritmo o gesticulación. Por su parte, Schneider y Preckel (2017) abordan esta noción en el marco de su dimensión “Presentación”, asignándole criterios de claridad y comprensibilidad, el estímulo del interés estudiantil y las habilidades de oratoria.

Este tipo de prácticas asociadas a las competencias y habilidades del docente para transmitir los conocimientos durante un curso, también ocupan posiciones intermedias en la escala de relevancia identificada por autores como Smith y Baik (2019). Por ejemplo, la riqueza episódica ocupa la décima posición en la lista. Por su parte, a la capacidad de comunicación (Asún y Zúñiga, 2017) en literatura de formación por competencias, se le entrega un bajo nivel de énfasis, la revisión de instrumentos previos un nivel medio y el estudiantado le asigna un alto nivel de relevancia. Estas observaciones son consistentes con los aportes de Schneider y Preckel (2017), que le asignan el cuarto lugar en términos de tamaño de efecto a la variable “presentación”. En otras palabras, de acuerdo con estos resultados, son los estudiantes quienes más valoran las capacidades comunicativas de un docente.

d. Dimensión de habilidades para la gestión docente

La **cuarta dimensión** corresponde a competencias que un docente debe desarrollar y ejercitar fuera del aula, pero que son fundamentales para la adecuada gestión y desarrollo de un curso.

- i. **Manejo de tecnologías:** Entendido por Schneider y Preckel (2017) como el uso apropiado de información y tecnologías de la comunicación, como clases online, podcasts y cursos online masivos (MOOCs). Esta dimensión no está presente en ninguno de los otros trabajos mencionados en este apartado del presente reporte.
- ii. **Responsabilidades administrativas:** Definido por Asún y Zúñiga (2017) como el cumplimiento de las funciones necesarias para el correcto desarrollo de la docencia, como asistir responsablemente a clases, recuperar las clases perdidas y respetar los horarios y plazos establecidos en el curso, tanto en la realización de actividades como en la entrega de evaluaciones.

La dimensión de competencias para la gestión y desarrollo de un curso recibe, en general, un bajo nivel de énfasis por los instrumentos y actores presentados en la investigación de Asún y Zúñiga (2017), mientras que en Schneider y Preckel (2017) sólo se presenta el criterio de manejo de tecnologías, atribuyéndosele el penúltimo lugar en términos de tamaño del efecto y cimentando la dimensión como una de las menos relevantes observadas en la literatura. Sin embargo, estos artículos refieren únicamente a la docencia presencial, mientras que, en las investigaciones referidas a la docencia virtual, las competencias tecnológicas y la capacidad de gestionar apropiadamente los tiempos y estructuras administrativas de un curso ganan especial relevancia.

e. Dimensión de estrategias docentes

Finalmente, la quinta dimensión considera las posibles prácticas y estrategias docentes utilizadas estimular el aprendizaje de los estudiantes. Se trata de la dimensión que agrupa la mayor cantidad de criterios, contribuyendo a corroborar la tesis de Schneider y Preckel (2017) de que las variables instruccionales son las más relevantes al momento de identificar prácticas educativas de alto impacto. Todas estas prácticas son definidas por Smith y Baik (2019), mas no son desarrolladas en profundidad por los demás autores tratados en este reporte. Por su parte, Asún y Zúñiga (2017) les entregan una cobertura parcial bajo el marco de la dimensión “Capacidad de manejo de diversas metodologías y técnicas de enseñanza”, mientras que Schneider y Preckel (2017) no hacen referencia alguna.

- i. **Aprendizaje basado en indagación:** Definido como la participación de los estudiantes en actividades investigativas orientadas a la resolución de problemas (Smith & Baik, 2019).
- ii. **Aprendizaje activo:** Entendido como el involucramiento activo de los estudiantes en su proceso de aprendizaje, ya sea por estímulo curricular, docente o involucramiento personal del mismo (Smith & Baik, 2019).
- iii. **Resolución de problemas:** Definido intuitivamente por Smith y Baik (2019) como el aprendizaje basado en la resolución de problemas a través de la aplicación de los contenidos entregados en el marco del curso.

- iv. **Inter-teaching/flipping:** Consiste en una de las estrategias más habitualmente utilizadas por docentes, y se define como la práctica de entregar materiales y contenidos al estudiantado por adelantado, mientras que durante la clase se produce un espacio de discusión y exploración de ideas asociados a dichos materiales (Smith & Baik, 2019).
- v. **Aprendizaje colaborativo:** Comprendido como la práctica de facilitar el aprendizaje de los estudiantes a través del trabajo en equipo (Smith & Baik, 2019).
- vi. **Aplicación del contenido:** Finalmente, la aplicación del conocimiento por parte del estudiantado en ambientes progresivamente similares al mundo real, o presentando problemas y necesidades propias del mundo real (Smith & Baik, 2019).

Las estrategias de enseñanza identificadas por Smith y Baik (2019) ocupan posiciones variadas en la escala de identificación en la bibliografía. Por ejemplo, prácticas como el aprendizaje colaborativo y basado en problemas ocupan el tercer y cuarto lugar, respectivamente, mientras que estrategias más específicas como la narración y uso de recursos multimedia no cumplen con la condición de tener más de una referencia, por lo que pueden considerarse como menos relevantes en el campo de estudio.

Por su parte, Asún y Zúñiga (2017) observan que la literatura relacionada con la formación por competencias, así como los instrumentos previos que revisaron sistemáticamente le atribuyen una importancia alta al manejo de diversas metodologías de enseñanza, mientras que los estudiantes le asignan una importancia media. Los niveles de valoración diferidos entre estrategias, así como la valoración intermedia del estudiante por esta dimensión sugiere que se trata de un conjunto de variables cuya importancia y aplicación responde al contexto disciplinar de carreras específicas más que a la generalidad de la educación superior, es decir, no existe un listado de estrategias de enseñanza que, aplicada genéricamente, pueda garantizar docencia de calidad, sino más bien un catálogo de herramientas cuya aplicación presenta mayor o menor impacto dependiendo de factores como el clima de aula, la carrera y recursos institucionales, existiendo contextos donde la docencia tradicional-expositiva presenta un mayor nivel de logro en la enseñanza y aprendizaje.

En resumen, en esta sección revisamos las diferentes dimensiones y criterios específicos que involucran una docencia de calidad, lo que está basado en las investigaciones de Smith y Baik (2019), Schneider y Preckel (2017) y Asún y Zúñiga (2017). Los criterios fueron ordenados según su relevancia y efecto en la enseñanza.

Desde la literatura, se observó que uno de los criterios más relevantes para garantizar una docencia de calidad es la estructura del contenido del curso, seguido por el alineamiento para que los estudiantes cumplan sus logros y que se asimilen a las evaluaciones (Smith & Baik, 2019; Schneider y Preckel, 2017). Además, Schneider y Preckel (2017) y Asún y Zúñiga (2017) destacan la interacción en clases como un importante factor para la formación estudiantil. En contraste, el manejo de tecnologías y responsabilidades administrativas tiene un bajo nivel de importancia (Asún y Zúñiga, 2017; Schneider y Preckel, 2017). Ahora bien, cabe tener en cuenta que estas investigaciones no tienen en cuenta en particular el contexto de docencia virtual.

En lo que respecta a la siguiente sección, se revisarán criterios de evaluación docente que utilizan distintas Instituciones de Educación Superior de Chile y del mundo. Además, se dará cuenta de los niveles de logro o de desarrollo de competencias docentes que están asociados a los mismos.

1.2. Criterios de evaluación que utilizan las Instituciones de Educación Superior

En la sección anterior, revisamos distintos criterios de evaluación de la docencia, los que han sido relevados en tres trabajos de investigación (Smith y Baik, 2019; Schneider y Preckel, 2017; Asún y Zúñiga, 2017), y han medido el efecto de las buenas prácticas en la enseñanza. Asimismo, se profundizó en distintos aspectos propios a la docencia en contextos virtuales. En este apartado en cambio, se presentan competencias o criterios que se obtuvieron de la revisión de modelos de evaluación nacionales e internacionales, además se ahonda en otros aspectos que estos presentan. Dentro de Chile, se revisaron los siguientes documentos, tomados de internet, referidos a evaluación y desarrollo docente en educación superior:

- i. Modelo integral del desarrollo de la docencia de la Universidad Católica de Temuco (UCT).
- ii. Marco para la Buena docencia de la Universidad Diego Portales (UDP).
- iii. Perfil docente de la Universidad de Concepción (UDEC).
- iv. Programa de desarrollo docente y Marco de cualificación de la docencia universitaria de la Pontificia Universidad Católica de Valparaíso (PUCV).

Entre las experiencias internacionales observadas podemos encontrar:

- i. El marco de normas profesionales para la enseñanza y apoyo al aprendizaje en la educación superior de la organización Advance HE del Reino Unido.
- ii. Marco de criterios y normas de la enseñanza de las Universidades Australianas, que incluye a la Universidad de Australia Occidental, Murdoch, Curtin, Edith Cowan y Notre Dame.
- iii. Criterios de evaluación de los Profesores de Track A y Profesores de Formación de Track B de la Universidad de Chicago.

Además, se incluye el Marco normativo e instrumentos de evaluación propuesto por Han et al. (2023). Estos marcos, criterios o modelos se construyeron con base en una revisión bibliográfica de literatura, y de experiencias internacionales y nacionales que abordan la docencia de calidad.

A partir de esta revisión, tanto en organizaciones como en instituciones de educación superior, se observaron dimensiones que dan cuenta de las competencias o criterios que usan las instituciones para identificar qué es buena docencia o docencia de calidad, así como los niveles de logro asociados al cumplimiento de estos criterios. Esto incluye a los docentes independientemente de su nivel, desde el profesor novicio hasta el titular. Además, cabe tener en cuenta, que, en general, en las instituciones y organizaciones observadas, no se profundiza sobre sus criterios o estándares para cada nivel jerárquico docente en particular.

a. Dimensiones de logro en buena docencia definidas a través de marcos, modelos y/o criterios institucionales chilenos

A nivel nacional, los cuatro modelos de desarrollo docente comparten diversos criterios (ver tabla 1). Aquellas dimensiones de la docencia que son consideradas únicamente como parte o nivel de logro de otra dimensión son consideradas como “parciales”.

Tabla 1: Dimensiones de desempeño docente en universidades chilenas

Dimensión/Universidad	UDP	UCT	UDEC	PUCV
Experticia disciplinaria	SÍ	SÍ	SÍ	PARCIAL
Planificación de la enseñanza	SÍ	SÍ	SÍ	SÍ
Metodologías didácticas	SÍ	SÍ	PARCIAL	SÍ
Clima favorable para el aprendizaje y relación con estudiantes	SÍ	SÍ	NO	SÍ
Evaluación	SÍ	SÍ	SÍ	PARCIAL
Reflexión e indagación docente	SÍ	SÍ	PARCIAL	SÍ
Ética profesional	SÍ	SÍ	NO	PARCIAL
Gestión docente	SÍ	PARCIAL	PARCIAL	SÍ

Fuente: Elaboración propia, 2024.

Entre estas habilidades, cada una de las universidades nacionales observadas considera diversos niveles y/o indicadores de logro. Mientras que existe un consenso sobre los criterios, sólo la UCT asigna niveles de logro jerarquizados por nivel de desarrollo profesional.

- i. La **experticia disciplinaria**, se asocia con la identificación de nociones y teorías de la disciplina que son relevantes, entrega de información actualizada a estudiantes y uso de bibliografía pertinente. En la UCT; el primer nivel se asocia con evidenciar el vínculo entre disciplina y quehacer profesional en el proceso de enseñanza y aprendizaje. Posteriormente, debe mostrarse un dominio profundo de contenidos, sumado al vínculo ya mencionado. Finalmente, se suma a todo lo anterior la incorporación activa de estos elementos al quehacer docente.
- ii. En cuanto al **diseño y planificación del proceso de enseñanza-aprendizaje**, se observan indicadores de logro relacionados con la claridad de programas de curso, el ajuste con los contenidos, actividades y métodos de evaluación, la presencia de estrategias diversas y la capacidad de realizar ajustes, entre otros. En el caso de la UCT, los tres niveles son, en primer lugar, evidenciar comprensión de los aspectos que orientan el diseño de la enseñanza-aprendizaje, segundo, incorporar ejes clave del modelo educativo institucional. Finalmente, incorporar estos ejes con foco en la contribución al perfil de egreso estudiantil.
- iii. En cuanto a la presencia de **metodologías didácticas**, se asocia con la diversidad de estrategias de enseñanza, y se expresa en la proposición de actividades variadas en consideración de la complejidad de los contenidos y la consideración de los espacios de expresión estudiantil. Al respecto, la UCT reconoce un primer nivel asociado a la consideración de características psico-socioculturales del estudiantado. Un segundo nivel dice relación con la generación de mayores espacios de autonomía para el desarrollo del aprendizaje. Finalmente, se considera la transformación de procesos formativos desde un enfoque responsable, autónomo y participativo.
- iv. En la dimensión **interacción y relación social con el estudiantado para un clima favorable al aprendizaje**, se consideran criterios como la demostración de entusiasmo e interés por los contenidos enseñados, la positiva acogida de comentarios e inquietudes estudiantiles y el fomento a la comunicación. La UCT jerarquiza el cumplimiento de este criterio indicando que, en una primera etapa, el docente debe ser capaz de comunicarse eficaz y respetuosamente con estudiantes en distintos contextos formativos. Luego, el docente debe motivar los aprendizajes. Finalmente, el docente debe poder liderar espacios comunicativos y participativos.

- v. Con respecto al criterio asociado a la **calidad de la evaluación de aprendizajes**, considera indicadores como la diversidad de estrategias evaluativas, la oportuna definición y comunicación de criterios para evaluar, el uso de evaluación formativa y la adaptación de procesos evaluativos. La UCT nivela estos logros, indicando que el primer nivel consiste en implementar criterios evaluativos en el programa de curso, alinearse con resultados de aprendizaje, socializarlos y retroalimentarlos y, finalmente, usarlos en la generación de evidencias sistemáticas sobre el aprendizaje.
- vi. Luego están las prácticas reflexivas y la constante autoevaluación de las acciones y efectividad de las prácticas docentes, la apertura a la observación y retroalimentación e implementación de innovaciones docentes. En algunos modelos, esta dimensión también se asocia con la investigación e indagación en la práctica docente.
- vii. En cuanto a la **actuación ética**, se consideran como indicadores de cumplimiento la demostración de dedicación a la disciplina, con trabajo de calidad, riguroso y disciplinado, el reconocimiento de la relevancia de la honestidad académica y el respeto a la puntualidad y compromisos adquiridos. La UCT asocia, a un primer nivel, el compromiso con la formación estudiantil y el proyecto institucional. El segundo nivel considera la actitud inclusiva y coherencia personal con estos valores. Finalmente, se espera que el académico sea reconocido por inspirar prácticas de impacto en la vida universitaria.

b. Dimensiones de logro en buena docencia definidas a través de marcos, modelos y/o criterios institucionales internacionales

A nivel internacional, se revisaron tres marcos y criterios. En estos, se pudo encontrar una serie de dimensiones comunes que permiten identificar competencias o buenas prácticas docentes relevantes a la hora de realizar una evaluación del profesorado, las que se muestran en la Tabla 2 a continuación.

Tabla 2. Dimensiones de desempeño docente según marcos y criterios de instituciones internacionales.

Dimensiones	Marco de criterios y normas de la enseñanza de las Universidades Australianas (AUTCAS en inglés)	Criterios de evaluación de los Profesores de Track A y Profesores de Formación de Track B (Universidad de Chicago)	Marco de normas profesionales para la enseñanza y apoyo al aprendizaje en la educación superior (Advance HE)
Diseño y planificación de actividades de aprendizaje, incluyendo dominio de la materia y métodos pedagógicos	SÍ	SÍ	PARCIAL
Evaluación y retroalimentación a los estudiantes	SÍ	SÍ	SÍ

Desarrollo de un entorno eficaz en torno a la comunicación, interacción y apoyo al estudiante	SÍ	SÍ	SÍ
Evaluación de la práctica para la continuidad del desarrollo profesional	SÍ	SÍ	SÍ
Efectividad profesional y personal, que incluye la interacción con pares	SÍ	PARCIAL	NO
Investigación y desarrollo de un aprendizaje profesional para incorporarlos a la práctica	SÍ	NO	SÍ
Enseñanza y apoyo al aprendizaje de los estudiantes	SÍ	NO	SÍ
Uso adecuado de tecnologías de la información	NO	NO	SÍ

Fuente: Elaboración propia, 2024.

- i. Una primera dimensión observada en la Tabla 2, refiere al **diseño y planificación de actividades de aprendizaje**, que, según el Marco de las Universidades Australianas o AUTCAS, no solo incluye planificar, preparar las clases y programas, sino también dominar la materia, manejar métodos pedagógicos para enseñarla, así como la coordinación con otros asistentes o ayudantes de cátedra para las clases. Además, para los profesores de nivel intermedio y avanzado, significa: 1. Asumir un mayor liderazgo en el desarrollo y diseño de plan de estudios, así como a nivel de programa a nivel nacional e inclusive internacional. 2. Contribuir en la enseñanza a nivel nacional, lo que se puede ver reflejado en las nominaciones a premios. 3. Y finalmente, en orientar y apoyar a sus colegas en tareas asociadas al diseño y planificación.
- ii. Luego encontramos la dimensión de **evaluación y retroalimentación de los estudiantes** que es compartida por todos los marcos revisados. Refiere a realizar tareas de evaluación que estén alineadas con los resultados u objetivos del curso, así como a las oportunidades para la mejora y práctica continua de los estudiantes en su aprendizaje. Además, el AUTCAS señala que, para niveles más avanzados, esto supone: 1. Mayor liderazgo en la realización de la evaluación de cursos y titulaciones. 2. Coordinar, apoyar y supervisar la evaluación, estándares y retroalimentación de los estudiantes. 3. Mostrar compromiso y conocimiento apropiado de las prácticas de evaluación. 4. Tener habilidades para evaluar y calificar tesis de postgrado. En concreto, para los profesores titulares, se espera que establezcan políticas y estrategias para la supervisión y evaluación del aprendizaje, y realicen tutorías exitosas a individuos o equipos para mejorar su evaluación a través de estándares.
- iii. En tercer lugar, se encuentra el **desarrollo de un entorno eficaz de aprendizaje**. Este considera la interacción y comunicación entre los estudiantes y docentes, y supone que el instructor sea sensible a las necesidades de sus estudiantes, que sea respetuoso y que desarrolle estrategias que den cuenta de su compromiso por fomentar la equidad y diversidad dentro del aula. En este caso, tal como se señala en el AUTCAS, se espera que, en un nivel intermedio y avanzado, el docente también: 1. Desarrolle iniciativas e innovaciones para el apoyo a los estudiantes, generando asimismo un ambiente atractivo. 2. Cumpla funciones de liderazgo promoviendo prácticas eficaces que consideren la

diversidad cultural e igualdad de los estudiantes, tomando en cuenta sus necesidades especiales. Además, en el caso particular de los docentes titulares, también se espera que dirijan políticas y estrategias de apoyo a los estudiantes y desarrollen tutorías a otras personas para apoyar la diversidad y transición a la educación superior.

- iv. Una cuarta dimensión, refiere a la **evaluación de la práctica para la continuidad del desarrollo profesional**, que da cuenta de la competencia y crecimiento en la carrera y métodos de enseñanza. En este caso, el AUTCAS identifica que, para los niveles más avanzados, esto supone desarrollar un mayor liderazgo y contribuir en el desarrollo profesional de los demás, desarrollar tutorías y evaluación de colegas en la enseñanza, evidenciar sus competencias y su compromiso sostenido en el desarrollo profesional a través de certificaciones en docencia, entre otras buenas prácticas. Además, al igual que en las dimensiones antes abordadas, también se espera que, en el nivel más alto, el docente dirija políticas y estrategias organizativas para apoyar y promover la evaluación docente.
- v. En cuanto a la **efectividad profesional y personal**, refiere a una serie de cualidades en las que se espera que los docentes demuestren progreso. Esta dimensión está mayormente desarrollada en el marco de las Universidades Australianas, que identifican aspectos tales como: demostrar compromiso con el desarrollo profesional continuo, responder positivamente a oportunidades y nuevos enfoques, comunicarse eficazmente en contextos informales y formales, aplicación de prácticas éticas profesionales. Además, como aspectos personales, se espera que el docente enseñe con entusiasmo, pasión y confianza, demuestre resistencia y perseverancia antes los obstáculos, gestione su tiempo y trabajo, sea autorreflexivo respecto de sus prácticas y demuestre interés por los estudiantes.
- vi. En torno a la quinta dimensión, involucra la **investigación y desarrollo del aprendizaje profesional**. En este caso, se pretende identificar que el docente desarrolle investigación para mejorar su propia práctica como la de otros docentes. Además, según se señala en el AUTCAS, para niveles más avanzados, esto significa mostrar un mayor liderazgo y contribuir eficazmente a nivel internacional en el desarrollo profesional y compromiso disciplinario, recibir reconocimiento de colegas, solicitar con éxito premios, becas o financiamiento para la enseñanza y aprendizaje, apoyar a colegas de niveles iniciales, y demostrar su autoría o co-autoría en publicaciones relacionadas al tema de enseñanza y aprendizaje.
- vii. Respecto a la **dimensión de enseñanza y apoyo al aprendizaje del estudiante**, considera el apoyo al aprendizaje de los estudiantes, entorno, asignatura y niveles de estudios. Como se observa en el AUTCAS, para esto es relevante que el docente sea creativo e innove para que de este modo logre motivar a sus estudiantes a aprender. Además, para los niveles más avanzados, se espera que el docente reciba un mayor reconocimiento de la calidad de su enseñanza, evidencie un desarrollo sistemático e integrado de prácticas basadas en evidencia, asuma mayor liderazgo e innove en prácticas de apoyo a los estudiantes, así como a colegas mediante la revisión. Particularmente en los docentes titulares, incluye un liderazgo universitario, disciplinario o internacional, y establecer políticas eficaces para promover y apoyar a otros para una enseñanza de calidad.
- viii. Finalmente, se encuentra el **uso adecuado de tecnologías de la información**. Esta competencia solo se identifica en el Marco de normas profesionales para la enseñanza y apoyo al aprendizaje en la educación superior del Reino Unido, y está considerada como

parte de los conocimientos fundamentales con los que debería contar un docente. En este caso, de manera similar a lo que se observa en AUTCAS, se espera que, en los niveles más avanzados, el docente desarrolle y aplique estrategias, políticas, procedimientos o iniciativas eficaces e integradoras para mejorar las prácticas y los resultados de los alumnos.

A partir de este apartado, pudimos revisar los distintos criterios de evaluación docente que se utilizan a nivel nacional, así como internacional. Así como el nivel de logro asociado a cada uno de ellos. En la siguiente sección, a modo de síntesis, se realizará una comparación general de los criterios de buena docencia que fueron identificados a través de los estudios de Smith y Baik, 2019; Schneider y Preckel, 2017; Asún y Zúñiga, 2017, así como desde la revisión de los modelos o marcos que han desarrollado distintas IES nacionales e internacionales.

1.3. Comparación general de los criterios de buena docencia

En la sección anterior, se presentaron las competencias o criterios que se obtuvieron a partir de una revisión de modelos de evaluación docente nacionales e internacionales, así como sus dimensiones de logro.

En este apartado, se busca hacer una comparación de los marcos, criterios o modelos nacionales e internacionales, así como los que se sostienen en la literatura para hacer una síntesis de los aspectos comunes que estos compartían. Esto también incluyó el marco normativo e instrumentos de evaluación propuestos por Han et al. (2023). A continuación, se señalan los criterios comunes y su respectiva explicación:

- i. **Dominio disciplinar y de habilidades pedagógicas para impartir el contenido curricular:** Se espera que el docente posea un dominio profundo y actualizado de su disciplina que lo conduzca a la innovación continua y la excelencia en su enseñanza. Además, debe tener un buen manejo de habilidades pedagógicas para enseñar el contenido a sus estudiantes. En el caso de la literatura revisada, esto se vincula con la expertiz en el área (Smith y Baik, 2019; Asún y Zúñiga, 2017).
- ii. **Diseño y planificación del proceso de aprendizaje:** el docente desarrolla actividades de aprendizaje que están planificadas y diseñadas teniendo como centro al estudiante. Además, los objetivos de aprendizaje y el programa del curso están claramente definidos, son atractivos y pertinentes y están dirigidos a los resultados.
- iii. **Evaluación y retroalimentación a los estudiantes:** el docente desarrolla diversas tareas de evaluación para medir los resultados de aprendizaje previstos, los que están alineados con los objetivos del curso. Además, ayuda a que los estudiantes puedan desarrollar y demostrar su aprendizaje, a través de una retroalimentación oportuna y clara
- iv. **Interacción y comunicación eficaz con los estudiantes:** el docente desarrolla un entorno eficaz, de apoyo y orientación al estudiante, para lo cual es importante que los escuche de forma empática y respetuosa, para lograr su atención y motivación durante el desarrollo del curso. Para ello, también debe crear vías de comunicación claras y coherentes (por ejemplo, definiendo horarios de atención)

- v. **Actuación ética y compromiso docente con el aprendizaje:** el docente muestra un compromiso con la formación personal y profesional de sus estudiantes, y la propia, actúa éticamente, responsable e inclusivamente considerando los principios y valores de la institución.
- vi. **Estrategias de enseñanza que se centran en el estudiante y sus especificidades:** el docente desarrolla estrategias de enseñanza considerando no solo su asignatura y el nivel de estudio en que se encuentran los estudiantes, sino también especificidades como la competencia, saberes y características psico-socioculturales que estos tienen.
- vii. **Evaluar la propia práctica docente para dar continuidad al desarrollo profesional:** el docente evalúa su propia práctica e implementa los cambios necesarios para impactar sobre el aprendizaje de sus estudiantes. Con esto se logra un permanente desarrollo profesional, además, en este también pueden colaborar otros pares.

Por otro lado, se identificó que había dimensiones menos comunes entre los modelos y criterios que se sostienen desde la literatura, las que se señalan a continuación:

- i. **Uso adecuado de tecnologías digitales:** el docente incorpora adecuadamente las tecnologías de información y comunicación en sus procesos de enseñanza y aprendizaje.
- ii. **Integrar la investigación a la práctica docente:** el o la profesora investiga sobre docencia universitaria y la integra a sus prácticas docentes para así mejorar los procesos de aprendizaje y formación.
- iii. **La interacción con colegas, así como asistentes de la enseñanza (ayudantes):** El docente se comunica con colegas y otros profesionales de manera oportuna y profesional, respetando políticas y obligaciones administrativas definidas por la institución. Además, proporciona orientación e instrucciones eficaces a sus ayudantes de cátedra, y los ayuda en su desarrollo pedagógico.
- iv. **Estimular el aprendizaje significativo:** el profesor transfiere valores, esquemas para la interpretación y pautas de comportamiento relevantes para el desarrollo profesional del estudiante en su área disciplinar.
- v. **Responsabilidades administrativas:** el docente cumple con funciones básicas como asistir a clases, recupera clases, respetar horario, entrega evaluaciones, etc.
- vi. **Uso de estrategias y prácticas docentes específicas para estimular el aprendizaje:** el docente desarrolla distintas estrategias, como promover en aprendizaje basado en actividades de investigación, estimula el involucramiento activo del estudiantado en su aprendizaje, se enfoca en la resolución de problemas para aplicar el contenido, entrega materiales y contenidos de forma previa a la clase para después discutirlos en el aula, promueve el trabajo en equipo y busca instancias en las cuales los estudiantes puedan aplicar el contenido al mundo real.

A partir de esta sección se pudo realizar una síntesis de los principales criterios de evaluación docente que se han identificado como relevantes en al menos tres estudios de investigación (Smith y Baik, 2019; Schneider y Preckel, 2017; Asún y Zúñiga, 2017), así como en modelos que han desarrollado distintas IES nacionales e internacionales. En el siguiente apartado, en cambio, se ahonda en el caso particular de la docencia en contextos virtuales y sus criterios de evaluación propios.

1.4. Docencia en contextos virtuales

En el anterior apartado se mencionó el conjunto de criterios para evaluar la docencia de calidad según sus dimensiones e importancia, considerando principalmente el contexto presencial, esto considerando tanto lo que se aborda en la literatura, así como a partir de los modelos que han desarrollado distintas IES nacionales e internacionales. En esta sección, en cambio, se profundizará en la enseñanza en contextos virtuales y posibles criterios para evaluarla.

La docencia virtual requiere de estrategias diferentes para garantizar la participación del estudiantado en el proceso de aprendizaje (Kebritchi et al., 2016). Una de las principales diferencias observadas en las **dimensiones de estrategia docente presencial y virtual** corresponde a que, en la segunda, las estrategias tecnológicas son parte fundamental de las habilidades a desarrollar en el contexto de la docencia. La evaluación de esta dimensión es un asunto con el que se debe tener mayores precauciones, toda vez que el levantamiento de criterios que contribuyan a la obligatoriedad de determinados estilos, estrategias o métodos de enseñanza podría ser mal recibido por una fracción importante de los docentes de diversas instituciones educativas. Más que evaluar el nivel de ejecución de estrategias de aprendizaje específicas, sería pertinente garantizar institucionalmente que el docente haya recibido formación en uno o más métodos de aprendizaje, dándole herramientas para innovar, pero dejándole en libertad de acción sobre su uso.

Para identificar los criterios de calidad para la educación superior virtual, González et al. (2023) desarrollaron 151 entrevistas virtuales a docentes durante la pandemia, período caracterizado por -entre otras cosas- iniciar un período de Gran Transición Online (GOT, por sus siglas en inglés). A partir de esta iniciativa, pudieron identificar dimensiones clave para el desarrollo de la docencia virtual (pp. 14):

- i. **Habilidades digitales básicas:** En el contexto de clases online, se identifica que es relevante el desarrollo de destrezas digitales básicas, sobre todo considerando que los niveles de competencias pueden ser muy variados entre los profesores. Estas requieren tener en cuenta distintas vías de aprendizaje y formación en el área para asegurar la calidad docente.
- ii. **Aplicación pedagógica de tecnologías digitales:** Con la llegada de la GOT, docentes tuvieron que aprender a priorizar contenidos y desarrollar planificaciones más detalladas. Además, estrategias tradicionales de evaluación, como la memorización, dejan de ser viables, por lo cual, los docentes deben comenzar a involucrarse en actividades similares a la auténtica evaluación. Los investigadores indican que se requiere guía sistemática en diseños curriculares y de aprendizaje para seguir construyendo sobre estas experiencias. También recomiendan el entrenamiento técnico para el aprendizaje y arraigamiento de habilidades digitales, así como marcos pedagógicos para comprender el rol de estas tecnologías y su influencia en el aprendizaje estudiantil.
- iii. **Uso de tecnología para el desarrollo profesional continuo:** Durante la transición online, los docentes valoraron tanto el desarrollo profesional pedagógico como tecnológico, sin embargo, se vieron sobrecargados por los programas de formación en estas áreas. Así, resultará importante incorporar las lecciones aprendidas en la pandemia; como ampliar el rango de actividades y considerar asuntos asociados a la carga laboral que pueden afectar la labor docente.

- iv. **Habilidad de profundizar competencias digitales para estudiantes universitarios:** Mientras que el estudiantado hace uso intensivo de redes sociales y aplicaciones virtuales, se vieron menos preparados para utilizar este tipo de tecnologías para su aprendizaje. A este respecto, los investigadores recomiendan que los docentes se involucren activamente en la profundización de las competencias digitales de los estudiantes.
- v. **Competencias transversales:** Se observa que, durante la pandemia, fueron fundamentales las habilidades de generar ambientes de aprendizaje seguros, ser accesibles para el estudiantado y ser comprensivos. Hacia el fin de la pandemia, los autores hacen hincapié en la necesidad de desarrollo profesional que contribuya a profundizar y mantener las competencias adquiridas durante la pandemia.

Por su parte, Orsini et al. (2022) presentan un conjunto de criterios para evaluar la calidad de los cursos dictados en este formato, validados a través de la aplicación digital del método Delphi. Estos criterios pueden organizarse en las dimensiones de:

- i. **Resultados de aprendizaje, metas y requerimientos:** Considera la claridad de logros de aprendizaje, la presencia de alineamiento constructivo y la presentación oportuna de la planificación del curso ante estudiantes. Este aspecto no es inherente a la virtualidad, y puede ser asimilado fácilmente por métodos de evaluación que no hayan sido expresamente diseñados para un contexto de docencia online.
- ii. **Diseño instruccional:** Considera la correspondencia entre actividades y necesidades estudiantiles, la coherencia en el diseño de cursos, el reconocimiento de múltiples modalidades de aprendizaje, el manejo del tiempo y la distinción entre actividades sincrónicas y asincrónicas. Aunque el diseño instruccional también es una dimensión para considerar en la docencia presencial, las habilidades docentes requeridas para ejecutar esta dimensión en la docencia online son diferentes y están directamente mediadas por la cantidad de herramientas a las que el docente pueda acceder y manejar competentemente.
- iii. **Presentación del curso:** Concentra elementos como la presencia de una página principal con recursos e información clara, secciones de orientación virtual, navegación intuitiva e instrucciones claras, además del uso oportuno y ordenado de plataformas virtuales institucionales donde se alojen los contenidos. Esto puede ser rápidamente evaluado por cualquier par que cumpla criterios de experiencia en manejo y diseño de plataformas virtuales educativas, así como los propios estudiantes que utilizarán estos recursos.
- iv. **Involucramiento estudiantil:** Contempla monitoreo de involucramiento, comunicación oportuna, estímulo a la expresión de ideas propias y disponibilidad de recursos de aprendizaje opcionales y de profundización. Este es un elemento que se considera en todo formato de docencia, pero que encuentra circunstancias y obstáculos especiales en la docencia virtual, donde la capacidad de alcanzar a estudiantes de manera significativa se ve muy reducida.
- v. **Uso de tecnologías para el estudiante:** Como observar la compatibilidad de los materiales con distintos tipos de dispositivos, mantener un inventario de recursos descargables, grabación de actividades sincrónicas y preocupación por formatos estándar y optimización de los materiales. Esta dimensión resulta especialmente pertinente para la docencia virtual,

y puede sintetizarse en el objetivo de hacer este formato más accesible para todos los estudiantes de un curso.

- vi. **Evaluación de módulo y curso:** Solicitar retroalimentación a los estudiantes, anonimizar dichos instrumentos, uso de metadatos de ambientes virtuales para informar la docencia y un plan de autoevaluación y mejora profunda.

Por último, está el estudio de Gómez-Rey et al. (2018), quienes también reconocen un conjunto de criterios que son relevantes para medir la buena docencia en contextos en línea, muchos de los cuales se alinean con los que identifican Smith y Baik (2019), Schneider y Preckel (2017) y Asún y Zúñiga (2017). En particular, entre las dimensiones mencionadas están las siguientes:

- i. Atributos pedagógicos del docente (considerando su pedagogía, el nivel de experticia en el contenido entregado y creación de materiales).
- ii. El diseño de las clases, que es fundamental para definir las tareas que se realizarán.
- iii. Las relaciones con sus estudiantes, especialmente en el contexto online donde es más difícil de gestionar.
- iv. La promoción de habilidades para la vida.
- v. El manejo técnico o tecnológico, que refiere a que el profesor sepa que ofrece la cultura digital en el ámbito educativo y lo utilice eficazmente.
- vi. Establecimiento de reglas, rutinas y procedimientos de aprendizaje.

En esta sección, pudimos profundizar en la enseñanza en contextos virtuales, la que requiere de estrategias diferentes para garantizar la participación de los estudiantes y por tanto criterios nuevos o adaptados para poder evaluar la calidad de la docencia.

Respecto a la docencia en línea, González et al. (2023), identifican cinco dimensiones claves para una buena enseñanza que surgen en el contexto de pandemia, estas refieren principalmente a habilidades que surgen de forma espontánea como resultado de la experiencia y que se recomienda seguir potenciando. Asimismo, Orsini et al. (2022), reconocen un total de seis criterios para evaluar la calidad de los cursos dictados en ese formato, algunos de los cuales se alinean con los que mencionan Smith y Baik (2019), Schneider y Preckel (2017) y Asún y Zúñiga (2017). Finalmente, están Gómez-Rey et al. (2018), que identifican al menos seis grandes dimensiones que se tienen que tener en cuenta a la hora de evaluar la buena docencia. Ahora bien, a diferencia de estos autores, González et al. (2023), Orsini et al. (2022) y Gómez-Rey et al. (2018), no identifican la relevancia de un criterio o dimensión por sobre otra. En este sentido, resulta relevante investigar cuál o cuáles son los aspectos que tienen un mayor efecto para asegurar así una enseñanza virtual de calidad.

En cuanto a la siguiente sección, se abordarán los diferentes instrumentos que se utilizan para evaluar la docencia, dando cuenta de sus aspectos positivos y negativos, así como tendencias emergentes que indican como estos han ido evolucionando en el tiempo.

2. Instrumentos para la Evaluación de la Docencia

En la sección anterior pudimos observar distintos criterios que se han utilizado para identificar que es una docencia de calidad y el nivel de importancia que tiene cada uno de ellos en base a lo que se ha planteado desde la literatura. Además, se presentó una revisión de modelos o marcos de competencias o criterios que utilizan distintas IES alrededor del mundo. Finalmente, se profundizó en las particularidades de la enseñanza virtual.

En esta sección, nos enfocaremos en los instrumentos que permiten observar el cumplimiento de los criterios que fueron mencionados en los apartados anteriores y que son importantes para la evaluación docente. Para cada uno de ellos se señala su definición, aspectos negativos y positivos y tendencias emergentes.

2.1. Cuestionario de Estudiantes Evaluando la Docencia (SET)

a. Definición

Según explican Spooren et al. (2013), el cuestionario de estudiantes evaluando enseñanza o SET (siglas en inglés para Student Evaluation of Teaching) habitualmente se desarrolla mediante la administración de encuestas estándar a los estudiantes, quienes deben rellenar ítems de calificación/opinión contruidos a partir de escalas Likert, además de algunas preguntas abiertas. El instrumento se suele aplicar de forma anónima, en formato presencial u online. Estos cuestionarios pretenden evaluar el rendimiento o calidad docente, además de evaluar un curso particular y otros posibles aspectos asociados. Asimismo, proporciona información para evaluar la promoción de un profesor y entrega evidencia de su responsabilidad con la institución (Chalmers & Hunt, 2016).

Si bien la mayoría de los estudios sobre el SET se han realizado en contextos presenciales, más recientemente el escenario online y las particularidades de esta modalidad de enseñanza han cobrado importancia (Gómez-Rey et al., 2018). En este contexto, se han realizado intentos por operacionalizar el concepto de buena docencia en el marco particular del aprendizaje en línea, con lo que se han desarrollado nuevos instrumentos SET. Un ejemplo, es el Gómez-Rey et al. (2018), quienes desarrollan un cuestionario a partir de serie de ítems/características que definen la enseñanza línea. Esta encuesta se construyó a partir de una amplia revisión de literatura, entrevistas a expertos y punto de vista de estudiantes a partir de un piloto del cuestionario.

b. Aspectos positivos

Entre las principales oportunidades que ofrece el SET según plantea Harrinson et al. (2022), se encuentra la facilidad de aplicarlo a diversas escalas y en distintos niveles disciplinares, además de ser un instrumento altamente replicable. Asimismo, representan un bajo costo en términos de recursos humanos, monetarios, y en cuanto al tiempo invertido en la recolección de datos y análisis, esto si se le compara con otros instrumentos, como la observación de pares, portafolio

docente y autorreflexión. También permite presentar resultados que se pueden cuantificar en porcentajes que den más claridad de ciertos aspectos que se quieren medir en cuanto al desempeño y calidad docente, relevantes para la promoción de cargos, renovación de los profesores, diseño y planificación de cursos, pero también para identificar y cambiar prácticas docentes dentro del aula.

c. Aspectos negativos

Una de las principales limitaciones del SET, es su carácter subjetivo, el que puede conducir a diferentes tipos de sesgos que afectan la fiabilidad y validez del instrumento. Una primera fuente que conduce a este sesgo es que estudiantes, profesores y administradores no siempre comparten un criterio común sobre buena enseñanza. Esto repercute en la comprensión de las preguntas, las respuestas entregadas, así como en la posterior interpretación de los resultados (Spooren et al., 2013).

Una segunda fuente sesgos que se debe considerar respecto al SET, se debe al estilo de respuesta, por ejemplo, la tendencia de estar totalmente de acuerdo con todos los ítems, o el sesgo de aquiescencia, que refiere a posicionarse en los extremos de la escala (Valencia, 2020). Esto conlleva a errores aditivos, es decir, cambios en la distribución de las puntuaciones, a la derecha o a la izquierda. Asimismo, errores correlacionales, los que cambian la relación con otras variables. Además, puede producirse una inflación de la calificación final y por tanto una dificultad o mala interpretación de los resultados por parte de los docentes. En este sentido, el autor nos recuerda que el SET no mide de forma directa la calidad de enseñanza, pues se infiere a partir de las respuestas brutas de los alumnos a la escala de valoración o los ítems disponibles (Valencia, 2020).

Finalmente, otro de los sesgos que se pueden producir en el SET, se relacionan con las características ya sea del estudiante, profesor o de quien lo aplica, las que no necesariamente están relacionadas con la calidad de la enseñanza (Spooren, 2021). Al respecto, Valencia (2020), señala como una de estas características, el género del docente. Pero además del género del docente, o incluso del estudiante, hay también otras variables que pueden estar afectando en las respuestas, tales como las calificaciones esperadas por los estudiantes, el tamaño de la clase (cantidad de alumnos/as), percepción de carga de trabajo del curso, atractivo del docente, la personalidad, si es un curso optativo o no, entre otras (Spooren, 2021).

d. Tendencias emergentes

Los comentarios abiertos en encuestas estudiantiles masivas, realizadas durante períodos semestrales de evaluación docente, son una valiosa fuente de retroalimentación para los profesores universitarios. Sin embargo, representa un desafío en cuanto a profundidad y escalabilidad para su análisis, debido al tiempo y esfuerzo necesarios para su procesamiento sistemático (Astudillo et al., 2023). Los comentarios estudiantiles sin filtrar pueden contener lenguaje ofensivo o insultos hacia el docente evaluado, lo que no solo es inapropiado, sino que también puede dañar la relación entre docente y proceso evaluativo y comprometer la credibilidad del instrumento. En este escenario, el avance en las tecnologías de inteligencia artificial ha ofrecido nuevas soluciones a este problema. Astudillo et al. (2023) aplicaron el modelo de inteligencia artificial Bidirectional Encoder Representations from Transformers (BERT) para analizar 4,335

comentarios de encuestas docentes de pregrado y posgrado en distintas especialidades de ingeniería en una universidad privada de alta selectividad. El modelo BERT, como su nombre indica, permite la representación vectorial del lenguaje natural, reconociendo el contexto de los fragmentos de texto y las relaciones semánticas entre distintas palabras.

La transformación del lenguaje natural a una representación vectorial es el primer paso para el análisis masivo de comentarios de evaluación docente usando el modelo BERT. Los autores definieron criterios de docencia de calidad basados principalmente en los trabajos de Schneider y Preckel (2017) y Smith y Baik (2021). En el modelo, estos criterios de calidad, simplificados lingüísticamente, se denominan "reactivos", los cuales también constituyen representaciones vectoriales. Posteriormente, se genera una tabla comparativa donde cada fila representa un reactivo o criterio de docencia de calidad, cada columna corresponde a un comentario estudiantil y las celdas contienen un coseno de similitud, es decir, la representación numérica de la asociación entre un comentario estudiantil y un criterio de docencia de calidad. Este método no solo permite el análisis sistemático de comentarios estudiantiles organizados según criterios de calidad definidos en la literatura, sino que también, debido al carácter cuantitativo de los datos obtenidos, facilita la aplicación de modelos de regresión y estrategias investigativas similares.

Sin embargo, existen limitaciones y desafíos en la aplicación de modelos como BERT. Las limitaciones institucionales en contextos como la educación superior chilena son notables, incluyendo la necesidad de computadoras con alta capacidad de procesamiento y personal con habilidades en análisis estadístico mediante plataformas como Python o R. Además, la calibración del modelo y la selección de datos requieren un riguroso proceso analítico.

2.2. Observación de Aula

a. Definición

Uno de los métodos para la evaluación de la docencia es la observación de clases. Este instrumento, ha sido ampliamente utilizado en educación para medir lo que hacen los estudiantes, lo que realizan los profesores y las relaciones que se producen entre ambos en el aula, con finalidad de evaluar el buen desempeño docente, así como las técnicas de enseñanza que está adoptando (Guimaraes y Lima, 2021). Este es llevado a cabo usualmente por docentes pares (Harrinson et al., 2020) o bien por personas capacitadas para realizar la evaluación.

Según plantean Guimaraes y Lima (2021) la observación, cuenta con diversos tipos y formas para registrar los datos, así como una gran variedad de protocolos que permiten gestionarla. A partir de una revisión sistemática de literatura publicada entre el 2000 y 2020 sobre prácticas de observación en contextos de aprendizaje activo en carreras STEM, los autores identifican en particular, cuatro grandes dimensiones que permiten agrupar los protocolos de ejecución:

- i. **Ambiente emocional e instruccional:** Contiene solo 3 protocolos de observación; dos de los cuales fueron aplicados en Estados Unidos. Consideró fuentes de datos cualitativas y cuantitativas.
- ii. **Manejo de la sala de clases:** También con 3 protocolos, todos aplicados en Estados Unidos.

- iii. **Evaluación de la enseñanza y el aprendizaje:** Es más extenso, contando con 18 protocolos de observación de clases, mayoritariamente desarrollados en Estados Unidos.
- iv. **Observación del comportamiento docente y estudiantil:** Es la más extensa de las categorías, concentrando 44 protocolos de observación de diversos contextos y nacionalidades.

De consecuencia, de dimensiones para la observación, pueden reconocerse múltiples sistemas y protocolos que permiten evaluar la docencia. Desde la segunda mitad del siglo XX, Borich et al. (1978) reconocían la existencia de diversos modelos, definidos como “instrumentos desarrollados para observar el comportamiento en clases de manera sistemática. Estos típicamente consisten en un número de categorías de comportamiento docente-estudiante, que el observador cuenta o califica periódicamente al observar las interacciones” (pp. 119). Además, los autores reconocían que diversos sistemas podían tener también diferentes grados de disponibilidad y uso en las comunidades educativas, distintos números de categorías y definiciones operacionales, así como grados de inferencia sobre las dinámicas que tienen lugar en el aula.

Más recientemente, Bell et al. (2019) identificaron las cualidades propias de cada sistema de observación a través de una revisión de cuatro protocolos. Para las autoras, existen categorías y aspectos necesarios para considerar en un **marco de evaluación de un método de observación**, como:

- i. **En términos de contenido;** las dimensiones de la docencia, la visión de la comunidad sobre enseñanza y aprendizaje, la especificidad del campo, “grain size” y foco en las acciones estudiantiles.
- i. **En términos de guía y orientación,** se requiere de procesos de puntuación definidos.
- i. **En términos de apoyo empírico,** se requiere de evidencia.
- i. **En términos de escala de implementación,** se necesita un continuo de desarrollo.

Para las autoras (Bell et al., 2019) la medición de la docencia a través de la observación requiere de procesos de muestreo de cursos (o partes de cursos), así como el aseguramiento de que las puntuaciones asignadas a dichos cursos son de calidad razonable. Esto implica la presencia de herramientas de puntuación, donde el sistema de observación determina cuáles serán las dimensiones medidas; sin embargo, la literatura ha demostrado que la observación humana tiende a ser poco confiable e imprecisa, especialmente cuando se consideran aspectos de la docencia como el desafío intelectual o la activación cognitiva, por lo que se sugiere la implementación de buenas prácticas como entrenamiento y certificación formal a través de evaluación a los observadores, como también doble codificaciones para evaluar la consistencia de las métricas.

Los sistemas de observación de clases también pueden ser adaptados a contextos de docencia virtual; Pusey et al. (2021) proponen el modelo E-COPUS, como adaptación virtualizada del sistema COPUS, basado en la identificación y categorización de las interacciones en el aula. En el caso de E-COPUS, algunas de las interacciones reconocidas son el uso de chats y la creación de grupos de trabajo en salas separadas, entre otros elementos.

b. Aspectos positivos

Entre las principales fortalezas que Harrison et al. (2020) identifica en la observación de pares, están el que estas instancias promueven las relaciones entre colegas, así como las discusiones y retroalimentaciones en temas de calidad de enseñanza y aprendizaje. Además, de acuerdo con los autores, esta estrategia tiende a ser más efectiva cuando sus procesos, tanto formativos como sumativos, son provistos por la propia institución, es decir, internamente.

c. Aspectos negativos

Por otra parte, como indican diversas investigaciones (Liu et al., 2019; Bell et al., 2018), el proceso de observación de los docentes no está exento de potenciales sesgos o errores de interpretación por parte de quien evalúa una clase, por lo que se requiere movilizar una importante cantidad de recursos para garantizar una adecuada aplicación, lo que implica dificultades para aumentar la escala en que este instrumento puede ser utilizado. Por ejemplo, la observación de aula requiere de entrenamiento y certificación para los evaluadores, diseños muestrales y sistemas de observación que permitan sistematizar los hitos que ocurren en la sala de clases.

Otras de las falencias de la observación, refieren a su enfoque en la *performance* del docente en un momento determinado de la clase. Esto, pues no considera que hay otras instancias relevantes que también influyen en el cambio de prácticas de enseñanza por parte del docente que ocurren fuera del aula y también surgen de la interacción con sus estudiantes (Esterhazy et al., 2021).

d. Tendencias emergentes

La diversidad de herramientas técnicas ofrece acceso a fuentes de información que anteriormente solo podían evaluarse intuitivamente o con un costo de tiempo y recursos que limitaba su escalabilidad. Un ejemplo es el análisis del desarrollo de la clase, incluyendo interacciones sociales entre docentes y estudiantes, estrategias de enseñanza aplicadas y competencias docentes. Aunque las clases pueden evaluarse mediante observación tradicional, esto requiere muchos recursos para evaluar una sola cátedra, limitando su aplicación sistemática (Harrison et al., 2020). En respuesta a este desafío, investigadores en la Iowa State University documentaron el uso del sistema TEACHActive. Este sistema utiliza cámaras en el aula para capturar clases, categorizando intervenciones estudiantiles e interacciones. Proporciona retroalimentación automática sobre las estrategias instruccionales, pero plantea desafíos éticos, como la falta de consentimiento explícito de los estudiantes y la reticencia de los docentes a ser evaluados mediante grabaciones (Kelley, 2021).

Debe considerarse que la estrategia de TEACHActive presenta diversos desafíos relativos al uso ético de los datos y la libertad de cátedra. Por ejemplo, no se solicitó consentimiento de los estudiantes para registrar las clases en que participaron, sólo se les informó de la iniciativa, toda vez que, de haber un solo estudiante que rechazara participar de la instancia, la grabación no podría haberse desarrollado (Kelley, 2021). Un segundo desafío corresponde al reclutamiento de docentes dispuestos a evaluarse a través de la grabación de sus clases, lo que ciertamente puede generar resquemores en comunidad académica.

En conclusión, podemos reconocer que los nuevos recursos técnicos proporcionan una gran variedad de datos para el análisis sistemático, pero que también implican desafíos en diversas líneas del quehacer evaluativo, por ejemplo:

- Requieren de recursos institucionales que no siempre son accesibles para las diversas instituciones de educación superior, como computadoras de alto rendimiento, servidores que alojen los datos, instalación de cámaras o incluso un soporte técnico permanente para garantizar su adecuado uso y funcionamiento.
- Se asocian a la demanda por funcionarios analistas altamente especializados en campos que han emergido recientemente, como el trabajo con modelos de inteligencia artificial.
- Representan desafíos éticos con respecto al uso responsable de los datos y de la información que de su análisis puede extraerse; recordando que tanto las técnicas de minería de datos como inteligencia artificial dependen de una rigurosa calibración humana.

2.3. Portafolio docente

a. Definición.

Los portafolios docentes han sido definidos como una colección de documentos y evidencia sobre el conocimiento, experiencia y habilidad docente (Wray, 2008). Según plantean Chalmers & Hunt (2016), este cuenta con al menos cuatro elementos para su elaboración:

- i. Una **declaración reflexiva sobre como enseñanza y por qué**. Qué valora del proceso de enseñanza y aprendizaje. Qué tipo de enseñanza aplica en el aula. Qué guía la toma de decisiones curriculares, y su evolución docente a lo largo de su carrera.
- ii. Una **descripción del contexto y objetivos institucionales e individuales**, una referencia a los valores y objetivos institucionales, la contribución de los profesores y la extensión, rango y duración de la enseñanza.
- iii. **Respuesta a los criterios institucionales de enseñanza** tales como diseño y aplicación de actividades, enseñar y apoyar a los estudiantes, evaluación y retroalimentación, desarrollo de entornos de aprendizaje eficaces, integración de actividades académicas, evaluación de la práctica y desarrollo profesional continuo, eficacia profesional.
- iv. **Liderazgo y resumen de los logros generales alcanzados** por los docentes, incluyendo las contribuciones en innovación.

En cuanto a la evaluación del portafolio, Chalmers y Hunt (2016, citados en Harrison et al., 2020) señalan que estos pueden ser evaluados tanto por revisores internos como externos, permitiendo la aplicación continua de un amplio rango de métodos para medir la calidad de la enseñanza.

Por otro lado, considerando el contexto de las clases online, así como el creciente uso de distintas tecnologías para impartir clases (Schneider & Preckel, 2017), el portafolio, como fuente documental o evidencia sobre el docente (Wray, 2008), también ha incorporado elementos que den cuenta de las habilidades docentes para el manejo de herramientas digitales y del cómo las utiliza durante las clases de manera concreta. Esto se incorpora como parte de su filosofía o

declaración reflexiva sobre su enseñanza, además puede ser ilustrado mediante ejemplos concretos de como incorpora la tecnología en sus clases.

A la construcción de un portafolio docente, puede sumarse un elemento autoevaluativo, entendido como una estrategia que estimula al docente a emitir juicios sobre su propio conocimiento, desempeño, creencias, productos y/o efectos en términos de efectividad. Este proceso implica la identificación de problemas, recopilación e interpretación de información y toma de decisiones (Schon, 1987, citado en Airisian y Gullickson, 1994). De acuerdo con Chalmers y Hunt (2016) la autoevaluación docente se relaciona directamente con la reflexión sobre la dinámica de enseñanza y aprendizaje; esta reflexividad resulta esencial para ir más allá de modelos tradicionales de docencia, como también contrarresta las visiones instrumentales y basadas exclusivamente en la competencia para evaluar a un docente. De esta manera, preguntas típicas de autoevaluación pueden ser: ¿Cuáles son mis responsabilidades profesionales sobre el aprendizaje estudiantil? ¿Cómo o en qué maneras soy responsable por mi docencia? ¿Qué responsabilidad tengo en mi crecimiento como docente? ¿Cuáles son los desafíos críticos de aprender mi disciplina?

A su vez, la autoevaluación implica recolectar, sistematizar y observar reflexivamente los materiales y estrategias utilizadas por el docente. Incorporar este elemento autorreflexivo sobre el propio quehacer y estilo de docencia puede ser especialmente enriquecedor cuando se contrasta con registros producidos por el diseño y desarrollo de un curso.

b. Aspectos positivos

Entre las ventajas que tiene este método según Fong et al. (2013) se incluyen que, a nivel individual, el portafolio digital de docencia ofrece un contexto para articular la propia filosofía docente, reflexionar sobre la propia docencia y documentar las pruebas de los logros. Esto constituye una fuente de datos de largo plazo que no solo permitiría mejorar las prácticas de enseñanza, sino que también podría ofrecer visiones detalladas sobre la calidad docente. Además, a nivel comunitario, el portafolio docente fomenta los diálogos sobre la enseñanza basados en la investigación, facilita el proceso de tutoría de los docentes novicios y ofrece recursos que ayudan a desarrollar criterios eficaces para la enseñanza en una institución terciaria. Por último, y según como se estructure la información, los portafolios pueden servir a distintos propósitos que deseen alcanzar los docentes (Wray, 2008).

Con respecto al elemento de la autoevaluación; Gómez y Valdés (2019) observan que esta contribuye a complementar y presentar una visión más holística de la calidad de la docencia, no sólo mejorando los procesos de evaluación, sino estimulando los procesos de auto reflexividad docente. Por su parte, Harrison et al. (2020) indican que esta forma de evaluación se constituye mecanismo de reflexión sobre la práctica docente. Su naturaleza autorreflexiva, bajo costo y alta replicabilidad lo convierten en una estrategia que puede contribuir a una discusión profunda sobre la calidad docente.

c. Aspectos negativos

En cuanto a los desafíos asociados con el uso de portafolios incluyen la intensiva movilización de recursos que requieren, la necesidad de evaluadores expertos para su revisión y la falta de anonimato del docente evaluado (Harrison et al., 2020). Otra de las falencias observadas por Berrill

& Addison (2010), es que usualmente para la evaluación de los portafolios no se utiliza una base o criterio compartido, ni se identifica el propósito principal para el cual se elabora, en consecuencia, las evaluaciones reflejan más las percepciones personales de quién los revisa.

En cuanto a la autoevaluación, Coverdale et al. (2009) expresan ciertas reservas, señalando su falta de una perspectiva integral y sugiriendo que la observación directa es el método más riguroso, mientras que la autoevaluación suele ser una fuente secundaria de información. Esto concuerda con Harrison et al. (2020), quienes relacionan esta estrategia más con la evaluación formativa que sumativa. Estos mismos desafíos convierten a la autoevaluación como un complemento de bajo costo y sumamente valioso para la evaluación docente a través de portafolios.

d. Tendencias emergentes

En los últimos años, el portafolio docente se ha trasladado a una versión digital, esto dada la emergencia de las tecnologías educativas en las instituciones terciarias de todo el mundo. Según señalan Fong et al. (2013), este tipo de portafolio puede liberar a los docentes a sus limitaciones geográficas y temporales, además parece ser una herramienta útil para promover el aprendizaje profesional. Sin embargo, también presenta retos asociados a las preocupaciones sobre el tiempo y recursos de apoyo necesarios para su elaboración que pueden repercutir en la intención de su uso.

Para concluir, en esta sección, pudimos conocer los distintos instrumentos que se han utilizado para evaluar la docencia en las instituciones de educación superior. En primer lugar, el cuestionario de estudiantes o SET, que utiliza como fuente de información la perspectiva de los estudiantes para evaluar el desempeño del docente. En segundo lugar, la observación del aula, que se realiza usualmente entre pares docentes que visitan la sala de clases para medir lo que hacen los estudiantes, lo que realizan los profesores y las relaciones que se producen entre ambos en el aula. Finalmente, están los portafolios, que son elaborados por los mismos docentes, y funcionan como una colección de documentos y evidencia sobre el conocimiento, experiencia y habilidad docente. Además, estos sirven para un proceso de autoevaluativo, que refiere a una estrategia que estimula al profesor a emitir juicios sobre su propio conocimiento, desempeño, creencias, productos y/o efectos en términos de efectividad.

Para cada instrumento, pudimos visualizar además aspectos positivos, negativos y tendencias emergentes que hay que tener en cuenta si queremos aplicarlo en nuestras instituciones. En la siguiente sección, se dan cuenta de ejemplos concretos de modelos de evaluación docente que se utilizan en distintas instituciones de educación superior a lo largo del mundo y que nos ayudarán a comprender cómo podríamos aplicar distintos métodos de evaluación en la práctica.

3. Ejemplos de Instrumentos de Evaluación Docente

En la sección anterior, pudimos conocer los distintos instrumentos que se han utilizado para realizar evaluación docente en el contexto de la educación superior, sus aspectos positivos, negativos y tendencias emergentes. Mientras que, en esta, veremos ejemplos concretos de cómo estos se utilizan en distintas instituciones de educación a lo largo del mundo. Esto nos ayudará a comprender cómo podríamos aplicar en la práctica estos distintos métodos.

a. Universidad de Melbourne

En la Universidad de Melbourne, existen al menos tres formas de evaluar a los docentes. La primera de ellas es la **revisión de pares de la facultad**. Esta pretende mejorar la enseñanza y aprendizaje en el aula. Todos los profesores pueden participar de estas instancias ya sea recibiendo o entregando retroalimentación y pueden escoger hacerlo voluntariamente. Para ello deben completar de forma previa un formulario con su información y planificar la revisión y **observación del aula**. Esto es intercambiado con sus colegas. Una vez coordinadas las fechas, se hace una observación durante las clases, para esto hay un marco o una base que guía la retroalimentación, de esta manera, solo se consideran aspectos relevantes para calificar la labor docente. Finalmente, se lleva a cabo la retroalimentación y discusión sobre lo observado.

La segunda forma de evaluación que se lleva a cabo en la Universidad de Melbourne es la **revisión formal de la enseñanza**, esta corresponde a una revisión holística del plan de estudios, prácticas de evaluación y enseñanza. La realizan expertos designados revisores por su trayectoria de excelencia docente, innovación y/o liderazgo. Este tipo de revisión, a diferencia de la revisión de pares, involucra presentar un **portafolio de enseñanza** que demuestre aspectos de la práctica y experiencia docente, enfoque de enseñanzas y ejemplos prácticos de sus propósitos educativos. También implica revisar el diseño curricular de la asignatura como requisitos y objetivos. Ahora bien, al igual que en la revisión de pares, se lleva a cabo **observación** durante la hora de clases para ver las interacciones entre docentes y estudiantes. Finalmente, con toda la información recogida mediante portafolio, diseño curricular y observación se debe llevar a cabo un informe escrito.

Finalmente, otra de las formas en que se evalúa a los docentes en Melbourne es a partir de la **Encuesta de Aprendizaje Estudiantil (SET)**. El cuestionario, considera la percepción de los estudiantes sobre sus experiencias de aprendizaje en determinadas materias, con el fin de mejorar la oferta educativa y la calidad de enseñanza. Estas se realizan al final del curso, y también a mitad del curso (a criterio del docente), es de carácter anónimo y permite responder preguntas específicas, generales antes de las evaluaciones finales. Si bien no es obligatorio responderlo en horario de clases, se incentiva a programar una hora para que los estudiantes la respondan y así mejorar la tasa de respuesta. Los resultados de la encuesta los reciben y revisan los docentes para actuar en base a ellos y mejorar así su enseñanza. Mientras que los estudiantes reciben comentarios de los resultados y las acciones que se han tomado al respecto.

b. Universidad de Carolina del Norte

En cuanto a la Universidad de Carolina Norte, cuenta con al menos tres instrumentos de evaluación. El primero, refiere a la **revisión de pares**, la cual se realiza explícitamente con fines formativos y sumativos. Esta involucra **observación de las clases** la que puede ser presencial u online. Para ello, se realiza una reunión de forma previa a la observación para clarificar qué aspectos revisar. Además de la observación, se realiza una **revisión del programa de estudios** que utiliza el docente y herramientas que usa durante las clases como páginas web, presentaciones, notas, textos, etc. El paso final, es la realización de un informe con una reflexión constructiva, es decir, que entregue comentarios que permitan mejorar la enseñanza. Además, para este tipo de evaluación, se asigna un revisor que tenga familiaridad con las prácticas de enseñanza propias de la disciplina, comprensión de la pedagogía, puede estar dentro o fuera del departamento, y de puede tener cualquier rango.

Una segunda forma de evaluar al docente en la Universidad de Carolina del Norte son los **cuestionarios a mitad de semestre**, los que, si bien son incentivados por la institución, son de carácter optativo. Estos incluyen generalmente preguntas abiertas para que los estudiantes comenten comportamientos y estrategias pedagógicas no cubiertas por las evaluaciones finales. Esta se desarrolla de manera anónima. Una vez recuperada la información, se buscan temas más comunes o que se repitieron para evaluar y mejorar.

La tercera y última forma de evaluación que utilizan en la Universidad de Carolina del Norte refiere al **cuestionario de evaluación de cursos** que se realiza a final de semestre a todos los estudiantes. Este es de carácter anónimo, y cuenta con preguntas cerradas y abiertas que están previamente designadas por la institución. Sin embargo, los profesores también pueden agregar sus propias preguntas. Los resultados si bien son visibles para todos los estudiantes, ellos no pueden aislar las calificaciones por instructor.

c. Universidad de Minnesota

En cuanto a la Universidad de Minnesota, utiliza dos instrumentos para realizar evaluación docente. El primero, corresponde a la **revisión de pares**. Este un proceso colaborativo donde el instructor de la clase o docente trabaja de cerca con un colega. Puede ser desarrollada con fines formativos o sumativos. Para llevarlo a cabo, se realiza una sesión de pre-observación, donde los colegas se ponen de acuerdo respecto de qué se va a observar, para luego llevar a cabo la observación en el aula, que tiene una duración aproximada de 50 minutos. En esta, el revisor puede o no participar de las clases y utiliza una rúbrica de evaluación. Después de la observación, se realiza una reunión donde se entrega feedback constructivo, un resumen de lo observado, incluyendo el contexto, fortalezas, planes de acciones y enfoque. Además de la observación también se hace una **revisión del programa del curso**. Este tipo de evaluación se realiza a todos los profesores titulares al menos cada dos años, y a los profesores en período de prueba una vez al año.

En cuanto al **cuestionario de estudiantes evaluando la enseñanza**, pretende entregar información sobre los docentes y cursos que imparten. Estas se realizan tres semanas antes del último día de clases, y se incentiva a que los profesores la realicen durante el horario de clase para mejorar la tasa de respuesta. Los resultados de esta encuesta se entregan a los docentes, además pueden

revisarlos sus pares, decanos o rectores de la universidad, pero no los estudiantes. Sin embargo, las respuestas respecto a determinados cursos si están disponibles para estos últimos.

d. Universidad de Stanford

La Universidad de Stanford utiliza distintas formas de evaluación docente, las que pueden aplicar los mismos docentes, por un revisor par, o a nivel institucional. Entre las iniciadas por el mismo docente, destaca la **retroalimentación de estudiantes durante la clase, y a mitad de semestre**. En esta participa un profesor que introduce la actividad y un entrevistador que la lleva a cabo. Para ello, se organiza la clase en grupos pequeños y realiza una serie de preguntas para mejorar la enseñanza y aprendizaje. En la actividad solo está presente el entrevistador, y el docente tiene que retirarse de la sala clases.

En segundo lugar, la Universidad de Stanford incentiva la **autorreflexión sistemática del docente**. Para ello es clave, elaborar un **portafolio** que entregue evidencia de su enseñanza, y declaraciones reflexivas sobre su práctica docente. También debe describir sus esfuerzos por mejorar su práctica o innovar en sus cursos, planes de estudio y tareas. Además del portafolio, es importante tener un inventario de su enseñanza, es decir, sobre cómo estructuran y llevan a cabo su clase.

Por otro lado, en la Universidad de Stanford, al igual que en las otras universidades, se desarrolla **evaluación de pares**. Para este tipo de revisión, se realiza inicialmente una **observación en el aula**, la cual es guiada en base a una pauta, que permite generar datos sistemáticos y comparables entre los colegas, por ejemplo, sobre el plan, materiales de estudios y objetivos de la clase. Además de la revisión, se evalúan los distintos materiales que utiliza el profesor para el curso, esto es, programas de estudio, tareas, exámenes, trabajos, etc.

Finalmente, están los **cuestionarios aplicados a nivel institucional**. Esta evaluación es realizada por los estudiantes, quienes hablan de su satisfacción con el curso y el desempeño de los docentes de manera individual. En específico, se evalúan aspectos del curso como objetivos de aprendizaje, elementos del curso y su organización. Aunque generalmente se realiza al final de período, puede completarse con una encuesta a mitad de semestre a criterio del docente. En cuanto a los resultados que se entregan, enfatizan la distribución de las respuestas más que resultados promedio.

e. Universidad de Oxford

La Universidad de Oxford incentiva a que los docentes utilicen al menos dos herramientas que les permitan evaluar y mejorar su enseñanza en clases. Una de las primeras prácticas, es la **retroalimentación durante la misma clase**. Para ello, los docentes pueden: 1. Consultar a sus estudiantes sobre aspectos positivos y negativos de las lecciones, recursos, tutoriales y otros materiales que se utilicen en el curso. Esto se puede hacer como una evaluación rápida en el aula o bien a través de sesiones de focus group donde se genere un espacio de diálogo. 2. Se pueden realizar formularios o herramientas de sondeo para recoger opiniones de los estudiantes. 3. Por último, elaborar cuestionarios breves para comprender experiencias de la clase semana a semana.

En cuanto a la segunda práctica que incentiva la Universidad de Oxford, son los cuestionarios de **evaluación durante el semestre**. Se aplican para calificar aspectos de la sesión de clases con

escalas de acuerdo, respuestas múltiples y preguntas abiertas. Estas son instrumentos elaborados por los mismos docentes.

f. Universidad de College London

La Universidad College London, al igual que la Universidad de Oxford, incentiva a que sus docentes utilicen distintas herramientas interactivas para evaluar su enseñanza durante las clases. Una primera forma, es a través de la **retroalimentación clase a clase**, donde los estudiantes pueden utilizar: 1. Sus propios dispositivos tecnológicos para entregar comentarios y hacer preguntas al docente 2. Notas adhesivas rojas, amarillas y verdes que evalúan las preferencias de aprendizaje y comprensión de los alumnos. 3. Papeles donde los estudiantes dediquen un minuto a escribir lo aprendido.

Una segunda forma en que los docentes pueden evaluar su enseñanza es a través de **cuestionarios a mitad de semestre** y que son aplicados a los estudiantes. En estos se prioriza lo que quiere saber el docente sobre sus estudiantes, y utiliza preguntas de escala de acuerdo, así como de respuestas abiertas, siendo similares a los cuestionarios de estudiantes evaluando la enseñanza o SET.

Finalmente, la University College London promueve la **autorreflexión de sus docentes** para la evaluación de enseñanza. Para ello, los profesores deben reflexionar sobre cada una de sus sesiones de enseñanza, y pueden utilizar herramientas para grabar las clases y así revisarlas después con detención.

Para cerrar, en este apartado pudimos revisar distintos instrumentos que utilizan en las instituciones para evaluar la docencia y la enseñanza, algunos de los cuales están más institucionalizados, como la **revisión de pares** que se puede aplicar en cualquier momento del semestre y **cuestionarios de estudiantes (SET)** que usualmente se aplican al finalizar el curso. Mientras que hay otros que, si bien se incentivan, dependen más bien de la iniciativa del propio docente para llevarlos a cabo, tal como la **retroalimentación de clases y autorreflexión**.

A continuación, se muestra una tabla resumen de cada uno de los instrumentos que se utilizan por cada Institución de Educación Superior (IES), el actor que evalúa, y tipo de evaluación: formativa y/o sumativa, donde la primera refiere a la evaluación durante el proceso de aprendizaje del estudiante y sumativa, una vez finaliza el proceso.

Tabla 3. Instrumentos de evaluación docente según IES.

IES	Instrumentos	Actor que evalúa	Tipo de evaluación
Universidad de Melbourne	Revisión de pares: Observación de clases	Pares Docente de la disciplina	Formativa.
	Revisión formal de enseñanza: Observación de clases y del programa de estudio	Pares Docente Expertos	Formativa.
	Portafolio docente		

	Cuestionario mitad y final de semestre	Estudiantes	Formativa y Sumativa.
Universidad de Carolina del Norte	Revisión de pares: Observación de clases y del programa de estudios	Pares Docentes de la disciplina	Sumativa y Formativa.
	Cuestionarios de mitad de semestre	Estudiantes	Formativa.
	Cuestionario final de semestre	Estudiantes	Sumativa.
Universidad de Minnesota	Revisión de pares: Observación de clases y del programa de estudios	Pares Docentes	Sumativa y Formativa.
	Cuestionario final de semestre	Estudiantes	Sumativa.
Universidad de Stanford	Retroalimentación de clases	Estudiantes	Formativa.
	Portafolio y Autoevaluación	El mismo docente	Formativa.
	Revisión de pares: Observación de clases y programa de estudios	Pares Docentes	Formativa.
	Cuestionario mitad y final de semestre	Estudiantes	Sumativa y formativa.
Universidad de Oxford	Retroalimentación de clases	Estudiantes.	Formativa.
	Cuestionario final de semestre	Estudiantes	Sumativa.
Universidad de College London	Retroalimentación de clases	Estudiantes	Formativa.
	Cuestionarios a mitad de semestre	Estudiantes	Formativa.
	Autorreflexión	El mismo docente	Formativa.

Fuente: Elaboración propia, 2024.

4. Un ejemplo de convergencia entre los criterios e instrumentos de evaluación docente: El Marco de las Universidades Australianas

En las secciones anteriores, se indagó en los criterios de buena docencia que son relevados desde la investigación, así como los que surgen desde los modelos de evaluación del profesorado que utilizan las IES. Además, se hizo una revisión de los distintos instrumentos de evaluación docente, sus aspectos negativos, positivos y tendencias emergentes, y se dieron ejemplos de cómo estos son implementados en distintas instituciones alrededor del mundo. En esta sección, se dará cuenta de cómo pueden converger ciertos criterios e instrumentos, para medir dimensiones específicas de la buena docencia, tomando en cuenta el Marco de las Universidades Australianas, que considera para cada dimensión al menos un instrumento de los que se han abordado en este documento.

El Marco de las Universidades Australianas, considera siete dimensiones para la evaluación docente. Para cada uno de ellos identifica el o los instrumentos que nos permitirían medirlo de forma óptima, tal como lo podemos observar en la Tabla 3, a continuación.

Tabla 4. Instrumentos según dimensión identificados en el Marco de criterios y normas de la enseñanza de las Universidades Australianas.

Dimensiones	SET	Revisión de pares, expertos o mentores	Portafolio
Diseño y planificación de actividades de aprendizaje, incluyendo dominio de la materia y métodos pedagógicos	Sí	Sí	
Evaluación y retroalimentación a los estudiantes	Sí		
Desarrollo de un entorno eficaz en torno a la comunicación, interacción y apoyo al estudiante	Sí		
Evaluación de la práctica para la continuidad del desarrollo profesional	Sí		Sí
Efectividad profesional y personal, que incluye la interacción con pares	Sí		
Investigación y desarrollo de un aprendizaje profesional para incorporarlos a la práctica		Sí	
Enseñanza y apoyo al aprendizaje de los estudiantes	Sí	Sí	

Fuente: elaboración propia a partir del Marco de criterios y normas de la enseñanza de las Universidades Australianas, 2024.

En general, de la Tabla 3, podemos observar que el Cuestionario de estudiantes evaluando la enseñanza o SET, es el que permite identificar el cumplimiento de la mayoría de las competencias

docentes, excepto por la investigación. En un segundo lugar, está la revisión de pares, relevante para revisar el diseño y planificación de la clase, la enseñanza y apoyo al aprendizaje de los estudiantes, así como la investigación para mejorar prácticas docentes - aspecto que no cubre el SET. Por último, encontramos el portafolio, que permite al docente hacer una evaluación de su propia práctica para la continuidad del desarrollo profesional. Esta última dimensión o competencia es una de las más importantes, pues permite al docente identificarse en los niveles que se encuentra, como profesora o profesor novicio, intermedio y titular, y así desarrollar prácticas para avanzar en su carrera profesional. Además, el AUTCAS señala indicadores relevantes para evaluar prácticas de buena docencia que están presentes en niveles más avanzados de la jerarquía docente. Entre los cuales se señala: recibir cartas de recomendación, recibir nominaciones a premios, invitación a conferencias y seminarios nacionales e internacionales, desarrollar supervisión a estudiantes de postgrado de forma exitosa, entre otras.

En esta sección, revisamos a grandes rasgos el Marco de criterios y normas de la enseñanza de las Universidades Australianas, el que nos permitió ver cómo pueden converger ciertos criterios de buena docencia con determinados instrumentos. Por ejemplo, la Evaluación y retroalimentación a los estudiantes se puede recoger de mejor manera desde el SET, mientras que la Evaluación de la práctica para la continuidad del desarrollo profesional, se alinea más con el portafolio. En la siguiente sección en cambio, se abordan las diferentes alternativas para implementar los diferentes instrumentos de evaluación que hemos revisado: SET, observación y portafolio docente.

5. Alternativas para la implementación de los Instrumentos de Evaluación Docente

En la sección anterior, pudimos ver a grandes rasgos cómo podemos hacer converger los criterios de evaluación docente con determinados instrumentos, tomando como ejemplo, el Marco de las Universidades Australianas. En esta sección en cambio, se señalan aspectos comunes y continuos que tienen los instrumentos de evaluación docente (SET, observación y portafolio), que son relevantes para el diseño y aplicación práctica de la evaluación del profesorado en las IES (Harrison et al., 2020; Chalmers y Hunt, 2016). Además, al final del documento de este apartado, se recomiendan ciertas alternativas de implementación basadas en la literatura abordada.

5.1. Características generales de los instrumentos de evaluación docente

Las elecciones respecto al diseño de los instrumentos de evaluación docente en general pueden responder a decisiones y factores contextuales propios de las disciplinas o facultades de las IES. Chalmers (2019, pp. 20) identifica dos continuos propios de la evaluación de pares docentes en el contexto australiano. El primero, es relativo a la afiliación de los revisores, que puede ser interna o externa a la institución. El segundo, refiere al propósito de la revisión, que puede ser sumativa o formativa. De acuerdo con la autora, existen beneficios para las distintas posiciones de estos continuos. Por ejemplo, identifica que la revisión externa es extensamente aceptada y practicada en la academia para propósitos sumativos, por su ser más proclive a la “independencia del juicio”; sin embargo, tiende a generar resistencias en los docentes. Por otra parte, la revisión interna es más apropiada para propósitos formativos por su grado de familiaridad entre revisores y docentes en evaluación, promoviendo un proceso más constructivo. La autora indica que “a pesar de haber un alineamiento natural entre la dimensión externa para propósitos sumativos e interna para resultados formativos” (pp. 19) existe una cantidad creciente de revisiones externas que persiguen un propósito formativo, así como modelos híbridos donde se mezclan actores externos e internos para una finalidad sumativa.



Un instrumento de evaluación docente puede generar distintos tipos de datos, que deben ser triangulados para presentar una visión holística del desempeño docente. Mientras instrumentos como SETs se asocian más directamente a la dimensión cuantitativa, que nos permite ver en porcentajes o cantidad qué aspectos de la docencia han sido mejor o peor evaluados por los estudiantes. Las estrategias como la observación de aula (Denaro et al., 2021) y la autoevaluación se pueden adoptar para generar información cualitativa que permita profundizar en aspectos específicos, como, por ejemplo, las prácticas metodológicas que utilizan en sus clases, la relación alumno y profesorado, etc.



5.2. Cuestionario de Estudiantes o SET.

El Cuestionario de Estudiantes o SET, puede aplicarse teniendo en cuenta el método y rigor científico. Esto implica, considerar aspectos durante su diseño, en el proceso de respuesta y en la interpretación de los resultados, para asegurar su validez y confiabilidad. En este sentido, un primer aspecto para considerar en la evaluación docente es que la elaboración de las preguntas de la encuesta esté basada en la teoría (Spooren et al., 2013). Esto es clave teniendo en cuenta que la docencia de calidad se refiere a un concepto abstracto, complejo y multidimensional, que requiere ser medido a partir de un gran conjunto de ítems, los que a su vez formen parte de una dimensión o constructo amplio, como, por ejemplo, la interacción social, habilidades de presentación, estrategias docentes, etc. Además, siguiendo el método científico, es importante verificar que exista una relación entre las preguntas que estamos realizando y el concepto de calidad docente, que es lo que realmente queremos medir (Valencia, 2020). Esto se conoce como la validez interna del instrumento.

En un segundo lugar, para que el puntaje o porcentaje obtenido en el SET tenga validez científica, es clave que la encuesta se haya testeado previamente. Por ejemplo, a través del juicio de expertos, que pueden ser tanto externos como internos a la institución, o también a través de un piloto de la encuesta que sea aplicado a un conjunto de estudiantes.

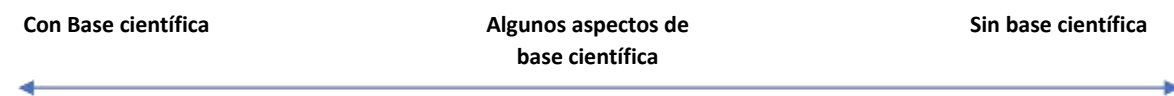
Un tercer elemento para considerar se refiere al proceso de respuesta del cuestionario, el que depende de la voluntad, el tipo de estudiante que responde, así como el contexto en el que lo hace (Spooren et al., 2013). Una forma de reducir las posibles diferencias en las respuestas debido a la incorrecta interpretación o formulación de preguntas por parte del estudiante es realizar entrevistas cognitivas, u ocupar la técnica de pensar en voz alta de forma previa a la aplicación de la encuesta.

Finalmente, otra forma de aseguramiento de la validez científica, pero menos directa, es ver la relación de las puntuaciones del SET con el rendimiento académico que han alcanzado los estudiantes en su carrera (Spooren et al., 2013). Ahora bien, también cabe considerar que los cuestionarios se aplican en distintos contextos disciplinares, por lo que también hay que ser cauteloso con el paradigma de aplicabilidad, y considerar así su validez en este nuevo entorno. Esto aún más si se ha traído desde otras instituciones, localidades o países (Spooren et al., 2013; Klemenčič & Chirikov, 2015).

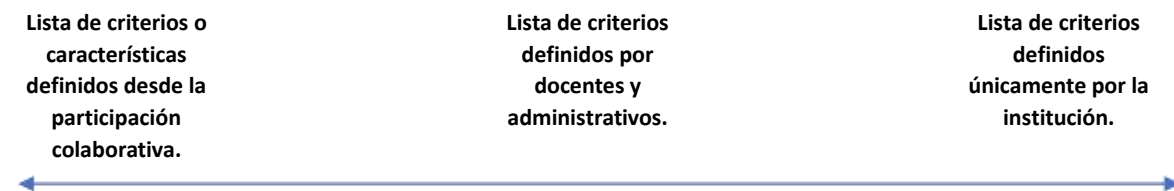
Si bien existen distintas técnicas para validar la puntuación entregada en la evaluación docente mediante el SET, como se mencionó en el segundo apartado de este documento, los resultados de la evaluación docente pueden presentar distintos tipos de sesgos. Entre ellos, se encuentran los errores por el estilo de respuesta (Valencia, 2020), sesgos asociados a las características que tienen los actores que responden o interpretan el SET, es decir, estudiantes, profesores o administrativos, las que no necesariamente están relacionadas con el desempeño docente (Spooren et al., 2021). Asimismo, variables tales como las calificaciones esperadas por los estudiantes, el tamaño de la

clase (cantidad de alumnos/as), percepción de carga de trabajo del curso, atractivo del docente, la personalidad, si es un curso optativo o no, entre otras (Spooren et al., 2021).

Por último, si bien es preferible aplicar el SET desde una base científica, también ello implica una gran inversión de tiempo y recursos humanos para lograrlo bajo la máxima rigurosidad. Considerando esto, también se puede llegar a una flexibilidad y equilibrio entre costo y beneficio al momento de aplicarla en las IES.



Según plantean Spooren et al. (2013), un claro entendimiento de lo que es una enseñanza efectiva es un pre-requisito para la construcción de un instrumento SET. Al respecto Penny (2003), propone establecer un grupo interinstitucional de consulta para establecer una lista de estándares o características dentro de un marco común de enseñanza efectiva, a lo que Spooren et al. (2013) añaden dos condiciones: 1. que las instituciones deben seleccionar aspectos más importantes de acuerdo con su visión y políticas, siendo consistentes con sus preferencias y 2. que todos los *stakeholders*, es decir, administradores, profesores y estudiantes deben estar envueltos en la definición de tales características. En la misma línea, Klemencic & Chirikov (2015), proponen que tanto los estudiantes, como otras partes interesadas deberían participar en el proceso de diseño, aplicación, análisis y elaboración de informes de las encuestas para contribuir en su pertinencia, calidad y legitimidad.



Si bien está bastante aceptado que el SET debe ser multidimensional, no está claro si se deben interpretar sus resultados como una puntuación global o por cada dimensión en particular. En ese sentido, un punto importante es distinguir el número y dimensiones de lo que es una enseñanza eficaz y la posibilidad de compilar todo lo anterior en un solo puntaje. El feedback y puntuación que entregan los estudiantes sobre aspectos particulares por una parte será útil para mejorar las prácticas docentes en un curso en particular. Al mismo tiempo, como el SET se usa para la toma de decisiones administrativas (sumativas), hay que considerarlo unidimensionalmente con un puntaje global que dé cuenta del desempeño docente. Como resultado, según plantean Spooren et al. (2013), existe un compromiso en el cual se recomienda el uso de dimensiones específicas para fines formativos, y medidas globales para fines sumativos, utilizando las medias ponderadas de las dimensiones individuales para generar una calificación global.



Según plantean Spoooren et al. (2013), es importante que a la hora de interpretar los datos del SET el docente no lo haga por sí solo, si no que reciba apoyo de pares, colegas y otros profesionales. En esta misma línea, es importante formar a los administrativos en teoría educativa y estadística, y en literatura actualizada del SET. Ahora bien, para lograr que la consulta de retroalimentación sea eficaz, debe ir más allá de la simple interpretación y asesoramiento para la mejora. Al respecto Penny & Coe (2004; como se cita en Spoooren et al., 2013), señalan ocho estrategias importantes: i. Participación activa de los profesores en el proceso de aprendizaje, ii. Uso de múltiples fuentes de información, iii. Interacción con los compañeros, iv. Tiempo suficiente para el diálogo y la interacción, v. Uso de las auto-calificaciones de los profesores, vi. Uso de información de retroalimentación de alta calidad, vii. examen de las concepciones de la enseñanza, y viii. Establecimiento de objetivos de mejora.



Actualmente, el SET se ha expandido más allá del dominio institucional a sitios web para calificar a los profesores, como RateMyProfessors.com. El problema de estos sitios es que están sujetos a un sesgo de autoselección, ya que la respuesta a estos cuestionarios es voluntaria. De este modo, suelen ingresar estudiantes interesados en calificar el desempeño del docente, ya sea porque no les gusta un profesor o les gusta la manera en que hace sus clases, teniendo consecuencias en la representatividad de la muestra, ya que no estaría compuesta por una diversidad de estudiantes, sino por un tipo particular de alumnado. En ese marco, los datos deben tomarse con cautela, y no deben utilizarse para fines sumativos. Sin embargo, Silva et al. (2008; como se cita en Spoooren et al., 2013) plantean que estos pueden ser similares a los cuestionarios tradicionales, que se refieren a aspectos de la enseñanza, personalidad y calidad global, pudiendo ser útiles para la formación de los docentes y mejora de sus prácticas.



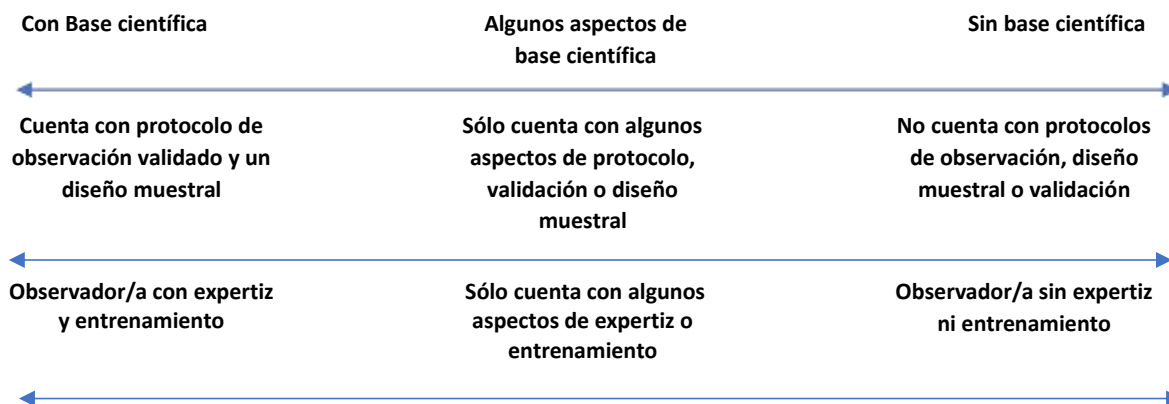
Según plantea Richardson (2005) y Klemencic & Chirikov (2015), lo más apropiado es que las encuestas de retroalimentación de los estudiantes se apliquen al finalizar la unidad del curso o programa de estudio para obtener información sobre la experiencia completa. Sin embargo, hay quienes señalan que obtener la retroalimentación al finalizar el curso no beneficia a los propios estudiantes, inclusive Greenwald & Gilmore (1997; como se cita en Richardson, 2005), señalan que realizar una encuesta a mitad de semestre sería más beneficioso pues influye en el estudio y calificaciones finales de los estudiantes a quienes se les aplica, permitiendo que se puedan generar cambios en las prácticas a tiempo. En cuanto a lo que se observa desde los ejemplos, en la mayoría

de las instituciones el SET se aplicaba finalizando el semestre, excepto en el caso de la Universidad de Melbourne, en el que se realiza también a mitad del curso, pero quedando a criterio del profesor.



5.3. Observación de Aula

Como el Cuestionario de Estudiantes o SET, la observación de clases requiere de diversos elementos para aplicarse adecuadamente y ser un aporte para evaluar la calidad docente. De esta manera, pueden identificarse ciertos continuos asociados a buenas prácticas, que contribuyen a garantizar la validez y confiabilidad del instrumento aplicado. En esta línea, Liu et al., (2019) argumentan sobre la relevancia de someter los sistemas de observación de clases a un proceso de validación en dimensiones asociadas a las formas y procesos de calificación, así como criterios de muestreo y puntajes asignados. Bell et al. (2018) coinciden con esta observación, además de identificar la necesidad de contar con observadores calificados, que hayan recibido entrenamiento y certificación. Finalmente, las autoras antes mencionadas coinciden en la necesidad de contar con protocolos estructurados de validación que permitan sistematizar los datos producidos en el marco de la actividad evaluativa de la docencia.



Además de los continuos antes mencionados, Liu et al. (2019) y Bell et al. (2018) reconocen algunos continuos de aplicación que no implican una diferencia en la calidad de los datos producidos, sino que se refieren a la adaptabilidad del instrumento a diversos contextos de docencia. Uno de estos ejemplos corresponde al foco de la observación, que puede estar centrado en las acciones de los docentes o de los estudiantes.

De acuerdo con Bell et al. (2018), el foco de aplicación de un instrumento de observación es altamente contextual, y depende de las escalas de clasificación, procedimientos y protocolos utilizados. En aquellos sistemas que hacen énfasis en el comportamiento estudiantil, se evalúan

acciones relativas a elementos afectivos, cognitivos y de comportamiento, como por ejemplo el ambiente de la clase, las dinámicas de trabajo conjunto, cuidado y respeto mutuo entre el estudiantado. Por otra parte, sistemas enfocados en los docentes se asocian más bien con prácticas para la enseñanza, el manejo de situaciones en el aula y la evaluación de aprendizajes, entre otros aspectos relevantes.



Por su parte, Bell et al. (2018) distinguen que un continuo o distintos grados de especificidad con que un sistema de observación aborda la docencia en educación superior, habiendo diseños contruidos para disciplinas específicas a diseños más generales, cuya aplicación es posible para cualquier carrera o entorno institucional. Ejemplo de esto es el modelo de Classroom Protocol For Undergraduate STEM (COPUS), diseñado específicamente para documentar observaciones de comportamiento en carreras de ciencia e ingeniería (Smith et al., 2013) a diferencia de protocolos como ICALT o CLASS K-3 UE, que son catalogados por Bell et al. (2018) como genéricos y, por tanto, de aplicación más abierta.



Una clase en educación superior se puede observar con fragmentos breves u holísticamente, observando todo su desarrollo. Bell et al. (2018) denominan este elemento como una dimensión de los procedimientos de puntuación, donde identifican sistemas que consideran ciclos acotados hasta la observación de una cátedra completa. Además, identifica sistemas como TIMSS, basados en la observación de videos de momentos específicos de la clase, que son abordados y analizados por equipos de diversos codificadores. Este elemento también se relaciona directamente con la definición de constructos a evaluar, toda vez que constituye la unidad de medida de la observación de clases (Liu et al., 2019).



El grado en que un instrumento contribuye a describir o interpretar la docencia se asocia con cuestiones de validez convergente (Borich, 1978) es decir, a la manera en que distintos instrumentos pueden medir un mismo constructo, o uno de carácter similar. De esta manera, entre la diversidad de instrumentos para la observación de clases podemos encontrar algunos que son altamente descriptivos, como COPUS (Smith et al., 2013) basados principalmente en la identificación y puntuación de acciones e interacciones reconocibles en el aula, como la cantidad de preguntas formuladas por estudiantes o el movimiento del docente por el aula. En contraposición, encontramos instrumentos enfocados en aspectos cognitivos o psicosociales de la

enseñanza y el aprendizaje, menos susceptibles de identificarse, describirse y puntuarse, precisamente por este carácter interpretativo con el que se asocian.

Aproximación descriptiva
de hechos y dinámicas

Aproximación
interpretativa



5.4. Portafolio docente

Para el portafolio docente, al igual que el SET y observación, también se considera un continuo relacionado con su base científica. Este tiene que ver con la elaboración de un diseño, proceso de respuesta y evaluación que permitan asegurar su validez (que mida lo que realmente pretende medir). En esta línea Torres & Zapata (2021), indican que un primer aspecto a considerar, son las distintas fuentes de evidencia que se utilizan para determinar los contenidos que deberían estar presentes en un portafolio, esto es, marcos legales que definen el proceso de evaluación docente, especificaciones de contenido en las que se basa la caracterización del desempeño docente, revisión bibliográfica actualizada, perspectivas de informantes clave sobre el sistema educativo y prácticas docentes, y finalmente datos auto informados sobre la autenticidad de las tareas del portafolio. Esto conforma un cuerpo de evidencias de contenido que da validez al uso del portafolio como evaluación docente.

En segundo lugar, Torres & Zapata (2021), señalan que hay que tener en cuenta el proceso de respuesta en el portafolio, refiriendo al momento en el cual el docente lo elabora. Para ello, es necesario entregarles instrucciones que evoquen efectivamente las prácticas y procesos cognitivos que se buscan evaluar, evitando respuestas sesgadas por problemas de claridad en la presentación de la tarea, o influencia de factores externos, como la falta de familiaridad. Para ello, se pueden realizar estudios piloto con otros profesores/as, y contar con datos relativos a las experiencias que tuvo el usuario cuando estaba completando el portafolio.

En tercer lugar, cabe considerar que los resultados del portafolio deben estar organizados en conjunto de categorías estándar distintas, bien definidas, finitas, exhaustivas y excluyentes para calificar el desempeño del docente. Para ello, los calificadores utilizan rúbricas definidas de antemano y emiten un juicio respecto a la calidad de respuesta basada en estas mismas. En este marco, se deben generar una serie de evidencias que den validez a la calidad de las rúbricas y del uso adecuado de estas, esto puede incluir la revisión de los aportes que ha hecho años anteriores y actualización de las pautas, estudios piloto con calificadores y estudiar procesos de respuesta en instancias grupales (Torres & Zapata, 2021).

Finalmente, los autores (Torres & Zapata, 2021) señalan que es importante tener en cuenta la validez interna del instrumento. Se asume que el portafolio mide un constructo que es el desempeño docente, en ese sentido, es importante analizar en qué medida la agregación de puntaje de cada indicador es consistente con la hipótesis de que hay un constructo único que nos permite medirlo. En este tipo de análisis, se observa en qué grado las relaciones entre los ítems se ajustan al constructo en que se basan las interpretaciones.

Con Base científica

Algunos aspectos de
base científica

Sin base científica

Según plantean Harrison et al. (2022), el portafolio puede ser una buena fuente de apoyo para que los docentes puedan documentar cómo han mejorado en sus prácticas de enseñanza después de haber sido evaluados en observaciones de pares o por sus estudiantes mediante un cuestionario. Sin embargo, como plantean Fong et al. (2014), este conjunto de evidencias podría ser útil para el asesoramiento de pares, pues facilita la observación de las clases y la autorreflexión del propio docente, como se observa en el ejemplo de la Universidad de Melbourne, donde el portafolio se incluye en la revisión formal de un conjunto de expertos a un profesor o profesora.

Fuente documental
de las prácticas
docentes que es
parte de la
observación de pares

Fuente para
documentar el
cambio en las
prácticas docentes
después de la
evaluación

Un aspecto de debate en torno al portafolio, como señala Buckridge (2008) es el propósito principal para el cual se realiza. Hay quienes señalan que el portafolio sirve para fines sumativos, como fuente de evidencia donde el docente puede dar cuenta de sus competencias y éxitos alcanzados. En este se haría una representación textual de la filosofía de enseñanza apoyada por logros y pruebas que corroboran dichas afirmaciones, teniendo como principal intención persuadir a otro. Esto, se usa para demostrar la calidad de enseñanza en un proceso de toma de decisiones como la titularidad y promoción de un profesor, y también para la entrega de premios y certificados.

En contraste con el carácter sumativo, otros señalan que el portafolio debiese cumplir fines formativos, enfocándose en que el docente presente su filosofía docente crítica y de cuenta de las competencias alcanzadas, reflexionando sobre sus prácticas docentes para mejorarlas y cambiarlas. Según plantean Leggett & Bunker (2006) esto ocurre principalmente en los portafolios para practicantes que recién están iniciando su carrera docente y en la cual trabajan en conjunto con un mentor quien presta mayor atención a los pequeños errores para que se puedan así corregir en la práctica. Ahora bien, igualmente Buckridge (2008) señala que, al largo plazo, la implementación continua de un portafolio enfocado en lo sumativo puede dar luces del desarrollo que ha tenido el profesor a lo largo del tiempo, conciliando así el énfasis en el logro o éxito, a la vez que la reflexión sobre los cambios que ha tenido en sus prácticas.

También relacionada al debate entre lo formativo y sumativo, está la cuestión de cómo se pide al docente que estructure la información del portafolio. Como plantean Chalmers & Hunt (2016), hay lineamientos o marcos que asisten a los profesores en el desarrollo de su portafolio, dando ejemplos de la práctica señalada por cada criterio, identificando las expectativas de rendimiento y recursos para evidenciar que se está cumpliendo. Esto sería clave tanto para que el docente pueda elaborar el portafolio, como en el momento de su revisión. Algunos autores criticaron estos estándares, pues si son demasiado rígidos, pueden llevar a que el docente responda como si

marcara una casilla de verificación, sin reflexionar sobre ellos. En ese sentido, se puede pensar que, en el extremo, se le dé mayor flexibilidad al docente al momento de escribirlo.



Otro de los aspectos importantes a la hora de evaluar un portafolio, tal como plantean Berrill & Addison (2010), es que de antemano el docente y revisor conozcan los criterios para su evaluación, así como los propósitos para los cuales se está realizando (entrega de premios, certificación, promoción, etc.). De otro modo, las evaluaciones serían más normativas, reflejando solo las percepciones personales del evaluador, sin guiarse por los criterios institucionales que se quieren visualizar a través del instrumento y los aspectos que se intentan medir en la rúbrica de evaluación.



Harrison et al. (2022) y Chalmers & Hunt (2016), señalan que, para la correcta implementación del portafolio en las instituciones de educación superior, es necesario contar con la suficiente cantidad de expertos en enseñanza y aprendizaje que puedan evaluarlos. Además, Chalmers & Hunt (2016), observan que este equipo, puede estar compuesto tanto por profesionales externos como externos a la institución. Sin embargo, si no se cuenta con esta disponibilidad, podrían participar un grupo de docentes pares que se ofrezcan voluntariamente a realizar esa labor, sin corroborar su preparación y capacitación previa.



Como estrategia complementaria, incluso cuando se le considera como un instrumento independiente, la autoevaluación implica la auro reflexión sobre registros de docencia e inventarios de prácticas educativas (Chalmers y Hunt, 2016). Así, la autoevaluación suele considerar pautas de evaluación guiada (Chalmers y Hunt, 2016), considera fuentes de información sobre la docencia aportando una dimensión autorreflexiva (Gómez y Valdés, 2019) y puede contribuir a calibrar otros instrumentos de evaluación (Roche y Marsh, 2000).

En esta sección, pudimos observar las diferentes alternativas para la implementación de los instrumentos de evaluación docente, que corresponden al SET, observación y portafolio. Con ello, se pudieron identificar que hay ciertos aspectos que han evidenciado ser más positivos en cada uno de ellos, en los que profundizaremos a continuación.

En cuanto al SET, se recomienda:

- i. Tener una **base científica y teórica** que dé confianza respecto de los resultados de los instrumentos. Esto conlleva a considerar la validez interna (relación entre los ítems o preguntas que componen el cuestionario en relación con la buena docencia), pasar por el juicio de otros expertos ya sea internos o externos, utilizar entrevistas cognitivas, la técnica de pensar en voz alta y otras estrategias de pre-testeo que aseguren la buena comprensión de las preguntas, así como verificar la relación entre los resultados del SET y el rendimiento académico. Además de considerar su aplicabilidad en distintos contextos y los sesgos que pueden producirse.
- ii. Contar con una **definición clara de lo qué es buena docencia**. Esto conlleva que todos los actores que intervienen en este proceso, es decir, estudiantes, docentes y administrativos, tengan un entendimiento en común de este concepto. Para ello, las instituciones pueden definir una lista con los criterios más importantes de acuerdo con su visión, pero además todos los stakeholders deberían estar envueltos en la definición de tales características. Esto puede hacerse desde el trabajo colaborativo.
- iii. Privilegiar la utilización del **SET para propósitos formativos**, que permitan identificar fácilmente la calificación que ha recibido el profesor para ciertos aspectos específicos que potencialmente se pueden mejorar. En otras palabras, se debe entender la buena enseñanza como un concepto multidimensional, que depende de distintos factores.
- iv. Considerar con el **apoyo de pares, colegas y profesionales para la interpretación** de los resultados. Esto conlleva que el docente participe en el aprendizaje, utilice distintas fuentes de información para comprender la calificación obtenida, cuente con tiempo suficiente para el diálogo e interactuar con sus colegas, se autoevalúe y haga un examen sobre sus concepciones sobre buena enseñanza y establezca objetivos para mejorarla.
- v. Contar con **cuestionarios que sean visibles y estén disponibles en el sitio web** del curso, para que las y los profesores puedan incentivar a sus estudiantes a responder. Esto es relevante, dada la baja tasa de respuesta de los cuestionarios online.
- vi. Realizar **encuestas de retroalimentación idealmente a mitad y final de semestre**. Al finalizar para obtener información sobre la experiencia completa y a mitad de semestre para que los docentes pueden identificar y generar cambios en sus prácticas a tiempo.

Respecto a la observación o revisión de pares, es recomendable:

- i. **Contar con sistemas de observación de clases, sometidos a validación** en las dimensiones asociadas a las formas y procesos de calificación del o la profesora, y en los criterios establecidos para el muestreo y puntajes asignados a los docentes. Contar con un grupo de calificadores que estén entrenados y que hayan recibido una certificación para realizar la evaluación. Así como tener protocolos estructurados para poder sistematizar los datos producidos.
- ii. En la aplicación, si el **enfoque está principalmente en el docente** que realiza la clase, es importante distinguir o identificar sus prácticas docentes, cómo maneja las situaciones en el aula y cómo evalúa el aprendizaje de sus estudiantes.
- iii. Para lograr un mayor grado de especificidad en el sistema de observación, se puede hacer un **diseño según el área disciplinar**. Esto permite ver aspectos particulares de las prácticas docentes que no se dan de la misma manera en todas las áreas, por ejemplo, en

humanidades, el enfoque puede estar en promover la reflexión y discusión de los estudiantes, mientras que, en ciencias naturales, la clase puede estar enfocada en que los estudiantes desarrollen tareas prácticas, por lo que el foco estará puesto en distintos aspectos.

- iv. Hacer una **observación holística de la clase**, es decir, de una cátedra completa, para reflejar la interrelación de las prácticas y capturar una imagen completa.

En tercer lugar, está el portafolio académico, el que idealmente debería:

- i. **Contar con una base científica y teórica** que dé confiabilidad al instrumento. En este caso, cabe considerar la validez de contenido del instrumento, que cuente con fuentes de evidencia, como marcos legales, bibliografía actualizada, datos autoinformados sobre la autenticidad de las tareas, etc. Segundo, que cuente con instrucciones que evoquen realmente las prácticas y procesos cognitivos a evaluar. En tercer lugar, el uso de rúbricas y su uso adecuado por parte de los calificadores. Finalmente, su validez interna, es decir, que existe una relación entre los ítems y que se ajustan al constructo, que, en este caso, es el desempeño docente.
- ii. Ser una **fuerza documental de las prácticas docentes** que apoye la interpretación de la observación de pares e interpretación de los resultados SET, de manera que se pueda tener una imagen mucho más completa del docente. También que apoye la autorreflexión sobre sus propias prácticas.
- iii. Desarrollar un **portafolio implementado de manera continua por el docente**, para poder dar cuenta del cambio en sus prácticas docentes. En este se hace una reflexión textual de la filosofía de enseñanza, así como logros y pruebas que la corroboren.
- iv. **Explicitar los criterios de evaluación y propósitos** para los cuales se desarrolla el portafolio de evaluación docente tanto por parte de los docentes como de los revisores.
- v. **Realizar la evaluación de este portafolio, a través de pares expertos** en temáticas de enseñanza y aprendizaje. Estos pueden ser externos e internos de la institución que estén previamente capacitados y dedicados a esta labor.
- vi. Finalmente, y de forma complementaria al portafolio docente, así como a la encuesta y observación de clases, la **autoevaluación** puede aportar ofreciendo una visión más holística sobre el quehacer del académico evaluado (Chalmers y Hunt, 2016), enriqueciendo así este sistema de evaluación docente. Esta autoevaluación debe considerar diversas fuentes de información, como registros sobre sus roles de liderazgo, mentorías y tutorías realizadas, estrategias educativas y evaluativas utilizadas, investigaciones e inventarios de prácticas (Chalmers y Hunt, 2016), las que después pueden ser incorporadas en el portafolio docente. Además, este puede ser un proceso guiado por la misma institución de educación superior, a través de pautas elaboradas que conduzcan la reflexión del docente sobre sus propias prácticas (Chalmers y Hunt, 2016).

En esta sección, pudimos ver las distintas alternativas para la implementación de los instrumentos de evaluación docente, es decir, el SET, observación de pares y portafolio docente. Para ello, se consideraron diferentes continuos, por ejemplo, aplicar el instrumento con base científica o no científica, con fines sumativos o formativos u otros

más específicos de cada uno de ellos. Además, se hizo un resumen de los aspectos que han evidenciado ser más positivos en cada uno de ellos.

En el siguiente apartado, en cambio, se habla de los actores que están involucrados en el proceso de evaluación docente. Estos son relevantes, en tanto cada uno de ellos, puede tener una perspectiva distinta de lo que es una docencia de calidad.

6. Actores involucrados en el proceso de evaluación docente

En la sección anterior, pudimos ver las distintas alternativas para la implementación de los Instrumentos de Evaluación Docente que hemos revisado, esto es, SET, observación de aula y portafolio. Además, se dieron ciertas recomendaciones basadas en la literatura abordada. En este apartado, abordamos otro aspecto importante dentro de este proceso: los actores que están involucrados en la evaluación.

Como hemos observado, los procesos de evaluación y desarrollo de la docencia son, idealmente, multidimensionales, y consideran las perspectivas y experiencias de estudiantes, pares evaluadores e incluso de los propios docentes en evaluación. Para Chalmers y Hunt (2016) las fuentes de evidencia para el análisis de la docencia consisten en la autoevaluación, la evaluación de pares y colegas, la perspectiva e indicadores de logro estudiantil. Por otra parte, Sorcinelli (2002) explora los centros de enseñanza y aprendizaje, entre cuyas funciones se encuentra el desarrollo docente, e indica la presencia de una diversidad de *stakeholders*, cuyas perspectivas deben ser incorporadas en acciones e intervenciones que apunten a la calidad de la docencia y un impacto significativo sobre el aprendizaje estudiantil.

En consecuencia, podemos identificar diferentes categorías de actores involucrados en un sistema de evaluación y desarrollo docente: estudiantado, pares académicos y docentes en evaluación, que a su vez pueden encontrarse en distintas etapas de desarrollo e inserción en el medio y, por lo mismo, requieren de acciones de acompañamiento y evaluación diferenciadas. Finalmente, podemos identificar también un nivel que considera las estrategias, modelos formativos y unidades de apoyo dispuestos por cada institución en consideración de los recursos disponibles.

6.1. Estudiantes

La perspectiva estudiantil es una de las principales fuentes de información que pueden utilizarse para la evaluación de la docencia superior (Chalmers y Hunt, 2016). Los estudiantes son quienes están en la mejor posición para juzgar aspectos como qué tan claro, interesante, respetuoso y justo es el profesor, y son los únicos que pueden decir cómo ellos influyen en su actitud hacia el curso y motivación por aprender (Felder & Brent, 2004).

Si bien las perspectivas de los estudiantes han evidenciado ser valiosas, esta fuente de información, así como el instrumento que se utiliza para recopilarla (SET principalmente), no han estado exentos de críticas. Estas apuntan por una parte a los sesgos estudiantiles (Kreitzer y Sweet-Cushman, 2021) que podrían tener su origen en la valoración de ciertas características que van más allá de la calidad de la docencia (Harrison et al., 2020). En segundo lugar, la heterogénea calidad de la

información recopilada, así como las competencias que el estudiantado pudiese tener para evaluar la calidad del quehacer docente (Hornstein, 2017), producto de lo cual, se ha recomendado que los instrumentos basados en la perspectiva estudiantil, especialmente aquellos de aplicación masiva en forma de cuestionarios, no se utilicen para evaluaciones sumativas, sino solo formativas. En tercer lugar, se ha observado que la participación estudiantil en instancias de evaluación docente se caracteriza por ser especialmente baja y con tendencia al desinterés (Hornstein, 2017). Por último, hay quienes argumentan que no es lo mismo medir la satisfacción de los estudiantes que la calidad de la docencia, pues son fenómenos que no necesariamente se relacionan entre sí.

En síntesis, la percepción estudiantil si bien aporta en gran medida a ver aspectos específicos del desempeño docente, que otros actores como los pares no pueden identificar, su interpretación debe ser medida y cuidadosa, considerando los sesgos asociados a características del docente, la falta de competencias para la evaluación de calidad y los bajos niveles de participación de estos actores, especialmente cuando se trata de instrumentos de aplicación masiva como SETs.

6.2. Pares y colegas evaluadores

Chalmers (2019), así como Felder & Brent (2004), identifican que los pares docentes corresponden a otra fuente de información importante. A diferencia de los estudiantes, ellos están mejores equipados para evaluar si un curso es adecuado y está actualizado, y si las tareas, exámenes y contenidos son coherentes con los objetivos del curso. Estos actores pueden ser externos como internos a las IES, y generalmente evalúan a sus pares docentes realizando una revisión de portafolios y/o mediante la observación de sus clases con un enfoque formativo o sumativo Chalmers (2019).

6.3. Docentes en desarrollo

En el sistema de desarrollo y evaluación docente, además de los estudiantes y pares, se debe también considerar a los mismos docentes y su estado de avance o nivel de jerarquía en la estructura académica. De acuerdo con Zacher (2019), la noción de desarrollo docente en contextos de educación superior considera tres niveles de inserción en la carrera académica: inicial, intermedio y tardío.

- i. En el **nivel inicial** pueden encontrarse estudiantes de doctorado y docentes que han ingresado de forma reciente a una Institución de Educación Superior. En esta etapa, él y la docente debiesen recibir formación básica y de carácter obligatorio en diversas temáticas asociadas a su rol. La mentoría recibida por académicos con más experiencia tiende a relacionarse favorablemente con el desarrollo de este tipo de docente (Tareef, 2013). Esta contribuye a fortalecer aspectos de docencia, investigación y planificación de carrera, siendo un elemento importante para alcanzar el éxito en la carrera académica (Fountain y Newcomber, 2016)
- ii. En los **niveles intermedio y avanzado**, se sugiere que deberían adoptarse mayores responsabilidades para con el desarrollo de otros académicos que recién ingresan a una Institución de Educación Superior, debido a la mayor experiencia y expertiz en temáticas de docencia en educación superior (Tareef (2013); Fountain y Newcomber, 2016). Así, en particular en el nivel intermedio, las y los profesores asociados, debiesen conservar elementos de formación en docencia, enfocados más bien en la reflexión e indagación

autónoma sobre estrategias e iniciativas y, al mismo tiempo, establecer vínculos con equipos de docencia e investigación integrados por docentes en niveles iniciales y mucho más avanzados.

- iii. Finalmente, en el **nivel avanzado**, compuesto por profesores titulares, se caracteriza por adoptar roles de liderazgo, mentoría de pares, iniciativas de discusión, investigación o formación en áreas del ejercicio docente relevantes para cada académico. De ese modo, implica labores de evaluación y acompañamiento de pares, articulación de iniciativas y mentorías a profesores asistentes que compartan áreas de interés disciplinar.

En esta sección, pudimos ver quiénes son los actores que intervienen en el proceso de evaluación de la docencia: estudiantes, principalmente a través del SET, pares docentes, a través de observación de aula, y los mismos docentes, quienes desarrollan su portafolio y pueden autoevaluarse. Además, vimos que los docentes pueden encontrarse en distintos niveles de desarrollo, lo que implica considerar distintas características a la hora de evaluarlo, por ejemplo, un profesor en un nivel avanzado, empieza a tomar roles de liderazgo, mentoría, investigación o formación.

Para dar cierre a este documento, a continuación, se entregan algunas reflexiones finales de lo que se ha abordado sobre los criterios de evaluación docente, instrumentos, alternativas de implementación y actores involucrados en este proceso.

7. Reflexiones finales

Como observamos en este documento, la evaluación y desarrollo docente constituyen un campo de acción en la educación superior que implica múltiples elementos: Criterios o estándares que indiquen la calidad docente, instrumentos que permitan medirlo y una serie de actores que están involucrados en el proceso.

El establecimiento de criterios o estándares de calidad docente cumple múltiples funciones: cuando se le observa desde una perspectiva evaluativa, permite fortalecer el aseguramiento de la calidad en educación superior a través de la medición del rendimiento de los procesos de enseñanza y aprendizaje. En cambio, cuando son vinculados a procesos formativos, adoptan un rol de perfeccionamiento docente, otorgándoles nuevas competencias para el desempeño de sus funciones. Por otra parte, cuando lo formativo y lo evaluativo del proceso se vinculan, nos encontramos con el imperativo de que los docentes de educación superior cumplan con los criterios y estándares de calidad definidos por los proyectos educativos de sus instituciones para poder acceder a nuevos estadios de su desarrollo. Un modelo de evaluación y desarrollo integral debe considerar la sinergia entre estándares, evaluación y formación.

En cuanto a los actores vinculados a los procesos de desarrollo y evaluación docente, también juegan un rol fundamental. Cuando este elemento es observado a la luz de los procesos evaluativos, podemos identificar las diversas formas en que éstos pueden contribuir a la producción de información; por ejemplo, hemos observado una discusión sobre la validez de la perspectiva estudiantil como método de evaluación docente sumativa (Hornstein, 2017) y explorado utilidades alternativas (Sporeen, 2013), como también la presencia de pares evaluadores (Liu et al., 2019) e incluso la perspectiva de la autoevaluación como estrategia para estimular la reflexión sobre el propio quehacer docente. De esta manera, la interacción entre los diversos actores de la educación superior y el proceso evaluativo presenta una diversidad de perspectivas sobre la docencia de calidad, la cual, como hemos visto, también implica distintos niveles de participación e incidencia en procesos sumativos.

Por otra parte, los mismos procesos evaluativos sumados a iniciativas de desarrollo docente influyen en el acceso a las distintas etapas de la carrera académica; esta diferenciación también hace necesario que los criterios se adapten a esos niveles. Por ejemplo, como se observó en el modelo de desarrollo docente de la Universidad Católica de Temuco, aunque los criterios y estándares de calidad son idénticos para todos los docentes de la institución, sus indicadores de logro son diferidos y organizados en tres niveles. De modo similar, el Marco de Criterios y Normas de la Enseñanza y Apoyo al Aprendizaje en la Educación Superior, elaborado por la organización de Reino Unido Advance HE, el Marco de criterios y normas de la enseñanza de las Universidades Australianas, así como el marco propuesto por Han et al. (2023) proponen una distinción jerarquizada por niveles. En segundo lugar, la interacción entre la formación docente y los diversos actores involucrados en el proceso de enseñanza y aprendizaje implica la conciencia sobre las necesidades estudiantiles y las perspectivas de pares formadores, fortaleciendo la reproducción de una docencia de calidad y buenas prácticas disciplinares.

De acuerdo con lo observado en el sistema de desarrollo docente de la Universidad Católica de Temuco, el Marco de Criterios y Normas de la Enseñanza y Apoyo al Aprendizaje en la Educación

Superior (Advance HE), el Marco de criterios y normas de la enseñanza de las Universidades Australianas y el marco propuesto por Han et al. (2023), puede profundizarse en algunos ejes orientadores para la adaptación de los criterios o estándares a distintos niveles de desarrollo docente:

- a. El nivel correspondiente a **docentes novicios** se asocia con el requisito de adquirir criterios y competencias necesarias para la docencia de calidad. De acuerdo con el marco de criterios de las Universidades Australianas, el docente de este nivel debe obtener una evaluación estudiantil, de puntuación media o superior durante dos años consecutivos en las siguientes dimensiones: diseño y planificación del aprendizaje, estrategias de enseñanza, evaluación y retroalimentación a estudiantes, interacción y comunicación eficaz con los estudiantes, así como en la evaluación de la práctica docente para dar continuidad a su desarrollo profesional. Ciertamente, esta etapa puede caracterizarse como una de adquisición y aprendizaje de competencias docentes básicas.
- b. El nivel que se vincula con **docentes asociados, o de posición intermedia**, busca un potenciamiento en el desarrollo a través de la innovación de las buenas prácticas, la reflexión e indagación sobre el quehacer docente. Según se señala en los marcos de criterios de las universidades australianas y del Reino Unido, los docentes deben asumir una mayor colaboración con sus pares en etapas iniciales de desarrollo, por ejemplo, mediante mentorías, asesoramiento o tutorías. Asimismo, según se señala en Han et al. (2023), en esta innovación en sus prácticas, juega un papel importante manejar las tecnologías de la información y comunicación (TICs) tanto en la interacción con sus estudiantes como en la impartición y planificar el curso.
- c. Finalmente, en una etapa de **desarrollo avanzado o titularidad** se espera del docente un potenciamiento en las buenas prácticas y la adopción de una posición de liderazgo con respecto a las buenas prácticas de enseñanza. Además, se espera que desarrolle políticas y estrategias para producir mejoras en la calidad docente en las distintas dimensiones abordadas. Han et al. (2023) señala que para los últimos niveles es fundamental el avance e innovación. Esto debería reflejarse en adquirir mayores competencias tales como: utilización y buen manejo de la TICs, reflexión sobre la enseñanza, la elaboración o diseño de contenidos didácticos como manuales y herramientas de apoyo que sirvan a sus pares, así como la gestión de equipos de enseñanza para definir prácticas de aprendizaje especializadas.

Por otro lado, teniendo en cuenta los elementos explorados previamente, se considera que un marco integral de evaluación y desarrollo docente puede favorecerse sustancialmente de incorporar los siguientes elementos por cada criterio de calidad docente que incorporen:

- a. **Indicadores de logro asociados a los niveles** de desarrollo de carrera académica establecidos por la institución. Por ejemplo, desagregando el criterio de “Dominio de estrategias y métodos de enseñanza” en tres niveles. Un nivel de adquisición de competencias para docentes en etapas iniciales, un nivel de reflexión e indagación para docentes en etapas intermedias y un nivel de liderazgo e innovación para docentes en etapas avanzadas.
- b. **Estrategias y programas institucionales de formación docente** que se orienten a desarrollar competencias y habilidades vinculadas a ese criterio específico. Por ejemplo,

nuevas metodologías docentes pueden ser aprendidas y adquiridas a través de seminarios y recursos informativos, mientras que criterios de experticia disciplinar pueden ser mejor abordados a través de mentorías de pares.

- c. **Alternativas para evaluar el criterio;** pues existen ciertas áreas de competencia docente que pueden evaluarse adecuadamente con estrategias específicas. Por ejemplo; criterios asociados con el diseño y planificación de actividades cursos y actividades formativas pueden ser abordados a través de una evaluación de portafolio docente, mientras que la observación de clases contribuye directamente a producir información sobre el desempeño en criterios como la construcción de ambientes propicios al aprendizaje y las habilidades comunicativas.

REFERENCIAS

- Airasian, P. W., & Gullickson, A. (1994). Teacher self-assessment: Potential and barriers. *Kappa Delta Pi Record*, 31(1), 6-10.
- Abello, D. M., Tapia, J. A., & Calderón, E. P. (2020). Development and validation of the Teaching Styles Inventory for Higher Education (TSIH). *Anales de psicología/annals of psychology*, 36(1), 143-154.
- Asún, R. A., & Zúñiga, C. (2017). Evaluación docente universitaria: Hacia una perspectiva unificada. *Revista de Sociología*, (32).
- Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School effectiveness and school improvement*, 30(1), 3-29.
- Berrill, D. P., & Addison, E. (2010). Repertoires of practice: Re-framing teaching portfolios. *Teaching and Teacher Education*, 26(5), 1178-1185.
- Borich, G. D., Malitz, D., & Kugle, C. L. (1978). Convergent and discriminant validity of five classroom observation systems: Testing a model. *Journal of Educational Psychology*, 70(2), 119.
- Buckridge, M. (2008). Teaching portfolios: Their role in teaching and learning policy. *International Journal for Academic Development*, 13(2), 117-127.
- Coverdale, John H., Nadia Ismail, Ayesha Mian, and Charlene Dewey. 2010. "Toolbox for Evaluating Residents as teachers" *Academic Psychiatry* 34 (4): 298-301.
- Chalmers, D. (2019). Recognising and rewarding teaching: Australian teaching criteria and standards and expert peer review. Department of Education and Training.
- Chalmers, D., & Gardiner, D. (2015). The measurement and impact of university teacher development programs. *Educar*, 51(1), 0053-80.
- Chalmers, D., & Hunt, L. (2016). Evaluation of teaching. *HERDSA Review of Higher Education*, 3, 25-55.
- Denaro, K., Sato, B., Harlow, A., Aebersold, A., & Verma, M. (2021). Comparison of cluster analysis methodologies for characterization of Classroom Observation Protocol for Undergraduate STEM (COPUS) data. *CBE—Life Sciences Education*, 20(1), ar3.
- Esterhazy, R., de Lange, T., Bastiansen, S., & Line, A. (2021). Moving Beyond Peer Review of Teaching: A Conceptual Framework for Collegial Faculty Development. *Review of Educational Research*, 91(2), 237–271. doi:10.3102/0034654321990721
- Felder, R. M., & Brent, R. (2004). How to evaluate teaching. *Chemical Engineering Education*, 38(3), 200-202.
- Fong, R. W. T., Lee, J. C. K., Chang, C. Y., Zhang, Z., Ngai, A. C. Y., & Lim, C. P. (2014). Digital teaching portfolio in higher education: Examining colleagues' perceptions to inform implementation strategies. *The Internet and Higher Education*, 20, 60-68.

- Fountain, J., & Newcomer, K. E. (2016). Developing and sustaining effective faculty mentoring programs. *Journal of Public Affairs Education*, 22(4), 483-506.
- González, C., Ponce, D., & Fernández, V. (2023). Teachers' experiences of teaching online during COVID-19: Implications for postpandemic professional development. *Educational technology research and development*, 71(1), 55-78.
- Gómez, L. F., & Valdés, M. G. (2019). La evaluación del desempeño docente en la educación superior. *Propósitos y representaciones*, 7(2), 479-515.
- Gómez-Rey, P., Fernández-Navarro, F., Barbera, E., & Carbonero-Ruz, M. (2018). Understanding student evaluations of teaching in online learning. *Assessment & Evaluation in Higher Education*, 43(8), 1272-1285. DOI: 10.1080/02602938.2018.1451483
- Guimarães, L. M., & Lima, R. D. S. (2021). A systematic literature review of classroom observation protocols and their adequacy for engineering education in active learning environments. *European journal of engineering education*, 46(6), 908-930.
- Harrison, R., Meyer, L., Rawsorne, P., Razee, H., Chitkara, U., Mears, S., & Balasooriya, C. (2022). Evaluating and enhancing quality in higher education teaching practice: a meta- review. *Studies in Higher Education*, 47(1), 80-96. doi:10.1080/03075079.2020.1730315
- Hinton H. Reliability and Validity of Student Evaluations: Testing Models Versus Survey Research Models. *PS: Political Science & Politics*. 1993;26(3):562-569. doi:10.2307/420006
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016.
- Han, X., Ge, W., Wang, S., Wang, S., & Zhou, Q. (2023). Standards Framework and Evaluation Instruments. In *Handbook of Teaching Competency Development in Higher Education* (pp. 37-62). Singapore: Springer Nature Singapore.
- Kebritchi, M., Lipschuetz, A., & Santiago, L. (2017). Issues and challenges for teaching successful online courses in higher education: A literature review. *Journal of Educational Technology Systems*, 46(1), 4-29.
- Kelley, J., AlZoubi, D., Gilbert, S. B., Baran, E., Karabulut-Ilgu, A., & Jiang, S. (2021, September). University Implementation of TEACHActive—An Automated Classroom Feedback System and Dashboard. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 375-379). Sage CA: Los Angeles, CA: SAGE Publications.
- Klemenčič, M., & Chirikov, I. (2015). How Do We Know How Students Experience Higher Education? On the Use of Student Surveys. En A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott, *The European higher education area: Between critical reflections and future policies* (págs. 361-379). Berlin: Springer Nature. doi:10.1007/978-3-319-20877-0
- Kreitzer, R.J., Sweet-Cushman, J. Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform. *J Acad Ethics* 20, 73–84 (2022). <https://doi.org/10.1007/s10805-021-09400-w>

- Leggett, M., & Bunker, A. (2006). Teaching portfolios and university culture. *Journal of Further and Higher Education*, 30(3), 269-282.
- Liu, S., Bell, C. A., Jones, N. D., & McCaffrey, D. F. (2019). Classroom observation systems in context: A case for the validation of observation systems. *Educational Assessment, Evaluation and Accountability*, 31, 61-95.
- Matthews, K., & Cook-Sather, A. (2021). Engaging Students as Partners in Assessment and Enhancement Processes. En M. Shah, J. T. Richardson, A. Pabel, & B. Oliver, *Assessing and Enhancing Student Experience in Higher Education* (págs. 107-124). Cham: Palgrave MacMillan. doi:10.1007/978-3-030-80889-1
- Ordoñez-Avila, R., Reyes, N. S., Meza, J., & Ventura, S. (2023). Data mining techniques for predicting teacher evaluation in higher education. A SYSTEMATIC literature review. *Heliyon*.
- O'Leary, M., Cui, V., & French, A. (2019). Understanding, recognising and rewarding teaching quality in higher education: an exploration of the impact and implications of the Teaching Excellence Framework.
- Orsini, C., Jennings, B., Rodrigues, V., & Tricio, J. (2023). The development of a reporting form for peer observation of online learning courses: An e-Delphi consensus study of educators working in health professions education. *European Journal of Dental Education*.
- Pusey, T. S., Valencia, A. P., Signorini, A., & Kranzfelder, P. (2021). Breakout Rooms, Polling, and Chat, Oh My! The Development and Validation of Online COPUS. *bioRxiv*, 2021-07.
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415.
- Roche, L. A., & Marsh, H. W. (2000). Multiple dimensions of university teacher self-concept. *Instructional Science*, 28, 439-468.
- Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin*, 143(6), 565.
- Smith, C. D., & Baik, C. (2021). High-impact teaching practices in higher education: a best evidence review. *Studies in Higher Education*, 46(8), 1696-1713.
- Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, 12(4), 618-627.
- Sorcinielli, M. D. (2002). Ten principles of good practice in creating and sustaining teaching and learning centers. A guide to faculty development: Practical advice, examples, and resources, 9-23.
- Spooren, P. (2021). Improving the Quality of Teaching. En M. Shhaah, J. T. Richardson, A. Pabel, & B. Oliver, *Assessing and Enhancing Student Experience in Higher Education* (págs. 159-190). Cham: Palgrave MacMillan. doi:10.1007/978-3-030-80889-1

- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research*, 83(4), 598–642. doi:10.3102/0034654313496870
- Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational research review*, 5(1), 25-49.
- Tareef, A. B. (2013). The relationship between mentoring and career development of higher education faculty members. *College Student Journal*, 47(4), 703-711.
- Torres Delgado, G., & Hernández-Gress, N. (2021). Research professors' self-assessment of competencies. *Future Internet*, 13(2), 41.
- Torres, D., & Zapata, A. (2019). The portfolio in the national teacher evaluation system in Chile: Collecting evidence of validity as part of the instrument construction process. In J. Manzi, M. R. García, & S. Taut (Eds.), *Validity of educational assessment in Chile and Latin America*. Ediciones UC.
- Valencia, E. (2020). Acquiescence, instructor's gender bias and validity of student evaluation of teaching. *Assessment & Evaluation in Higher Education*, 45(4), 483-495. doi:10.1080/02602938.2019.1666085

4º Piso Edificio Decanato de Educación UC
Campus San Joaquín.
Pontificia Universidad Católica de Chile.
Avda. Vicuña Mackenna 4860, Macul, Santiago, Chile.

www.nmedsup.cl



Síguenos en redes sociales:



@milenio_EdSup



@milenio_edsup1



www.linkedin.com/in/nucleomilenioedsup



Núcleo Milenio Educación Superior @nucleomilenioedsup