Cuadernillo técnico
de evaluación educativa

3

Definición del referente de la evaluación y desarrollo del marco de especificaciones

Centro UC
Medición - MIDE

Textos de divulgación

INEE
Instituto Nacional para la Evaluación de la Educación
México

Cuadernillo técnico
de evaluación educativa

3

Definición del referente de la evaluación y desarrollo del marco de especificaciones

Centro UC
Medición - MIDE

Textos de divulgación

INEE
Instituto Nacional para la Evaluación de la Educación
México

Definición del referente de la evaluación y desarrollo del marco de especificaciones

© Centro de Medición MIDE UC
Av. Vicuña Mackenna 4860
Macul, Santiago, Chile, cp 7820436

© Instituto Nacional para la Evaluación de la Educación INEE
Barranca del Muerto 341, col. San José Insurgentes,
Alcaldía Benito Juárez, Ciudad de México, cp 03900

Autoras
María Beatriz Rodríguez Frias, MIDE UC
María Paulina Flotts de los Hoyos, MIDE UC

Editora
María Rosa García González, MIDE UC

Corrección de estilo
Arturo Cosme Valadez, INEE
Lissette Sepúlveda Cepeda, MIDE UC

Coordinación General
Adriana Guadalupe Aragón Díaz, INEE
Marcela Cuevas Ossandón, MIDE UC
Marcela Ramírez Jordán, INEE

Diseño
www.iunta.cl

Índice

Presentación 1

Resumen 2

Introducción 3

Identificación del referente de la evaluación y definición del objeto de medida o constructo evaluado 5

La operacionalización del constructo mediante el marco de especificaciones 9

Los riesgos de evaluar sin especificaciones de contenido 10

¿Cuáles son los componentes de un marco de especificaciones? 13

La tabla de especificaciones como herramienta que enmarca el desarrollo del instrumento 14

La diversidad de tablas de especificaciones 15

Una guía de pasos para elaborar una tabla de especificaciones de una prueba 20

Consideraciones generales 21

Los pasos en la elaboración de una tabla de especificaciones 23

Criterios para verificar la calidad de una tabla de especificaciones 28

Consideraciones finales: ideas fuerza 30

Referencias 32

Presentación

El Instituto Nacional para la Evaluación de la Educación de México, INEE, y el Centro de Medición MIDE UC, de la Pontificia Universidad Católica de Chile, han gestado una colaboración para el desarrollo y fortalecimiento de capacidades en evaluación educativa, en profesionales del Instituto y de los equipos responsables de los Programas Estatales de Evaluación y Mejora Educativa (PROEME) y del Proyecto Nacional de Evaluación y Mejora Educativa de Escuelas Multigrado (PRONAEME), en el marco del Sistema Nacional de Evaluación Educativa (SNEE), en México.

El documento que a continuación presentamos constituye un material de consulta que forma parte de una serie de nueve cuadernillos, cuyo propósito es orientar la comprensión de los conceptos centrales de la medición y la evaluación educativas y su impacto en el diseño de instrumentos; considerando que el proceso evaluativo es una suma de decisiones que deben cuidar la coherencia de cada uno de los elementos y fases que lo componen.

Este material se ha organizado en una serie de cuadernillos con base en las siguientes temáticas:

1. Nociones básicas en medición y evaluación en el contexto educativo.
2. Confiabilidad, validez e imparcialidad en evaluación educativa.
3. Definición del marco de referencia de la evaluación.
4. Desarrollo de instrumentos de evaluación: pruebas.
5. Desarrollo de instrumentos de evaluación: cuestionarios.
6. Desarrollo de instrumentos de evaluación: pautas de observación.
7. Desarrollo de instrumentos de evaluación: tareas de desempeño y rúbricas.
8. Análisis y uso de resultados.
9. Uso de resultados y retroalimentación.

Esperamos que este material resulte de utilidad para los profesionales que se desempeñan en el contexto de la medición y evaluación educacional. En los cuadernillos encontrarán nociones y conceptos fundamentales, además de recomendaciones prácticas, y sugerencias bibliográficas para quienes deseen profundizar en cada una de las temáticas trabajadas.

Definición del referente de la evaluación y desarrollo del marco de especificaciones

Resumen

El cuadernillo que el lector tiene en sus manos pretende ofrecer una mirada sinóptica de una de las tareas iniciales, pero centrales, del proceso evaluativo. Inicial, pues antecede a las decisiones técnicas relacionadas con la metodología de evaluación y a las decisiones operativas del levantamiento de la información. Y central, pues de ella depende que el problema, fenómeno o constructo que se encuentra en estudio sea efectivamente abordado mediante la evaluación, y que el propósito que esta se plantea pueda ser cumplido.

Esta fase contempla dos complejas subtareas: la identificación de un marco de referencia que proporcione una noción clara del objeto de medida, su extensión y complejidad, y luego el desarrollo de un marco de especificaciones que oriente en concreto el diseño de la evaluación. Se analiza la complementariedad de ambas tareas y la importancia de ellas como argumentos de validez de contenido de una medición. Se presenta la diversidad de formatos que pueden poseer los marcos de especificaciones, siempre cumpliendo con el fin de operacionalizar el constructo evaluado y garantizar su adecuada representación en los instrumentos que serán diseñados a partir de ellos.

Finalmente, se propone un modelo de tabla de especificaciones, fundamentando su utilidad para el proceso de construcción de instrumentos, y se detallan los componentes que la conforman, así como los criterios que permiten verificar la calidad de su elaboración.

Introducción

La necesidad de explicitar el marco de referencia que orienta todo proceso evaluativo y de definir con claridad el objeto de medida o constructo que se desea conocer, es un requisito insoslayable para el diseño e implementación de evaluaciones válidas y confiables.

Llevar a cabo esta tarea, es el punto de partida para el objetivo de garantizar la coherencia en el tránsito por los distintos pasos implicados en el ciclo de la evaluación: el desarrollo de las metodologías y diseño de dispositivos para levantar la información; las definiciones sobre el modo en que esta será analizada y utilizada; y las decisiones orientadas a garantizar que el propósito perseguido con todos estos esfuerzos, sea finalmente alcanzado.

En el contexto de la mejora educativa, el propósito de la evaluación orienta y da sentido a la definición del objeto a evaluar, el que resulta necesario conocer en profundidad y con sistematicidad para informar ciertas decisiones o guiar ciertos planes de acción. En la mayoría de las evaluaciones educativas, este propósito refiere a constructos complejos o a dominios extensos de contenido, por lo que la sola declaración de este objetivo no garantiza una comprensión exhaustiva y unívoca del mismo. No será posible cumplir el propósito de una evaluación si la información que se levanta no corresponde, por definición o por extensión, al constructo o dominio que se explora.

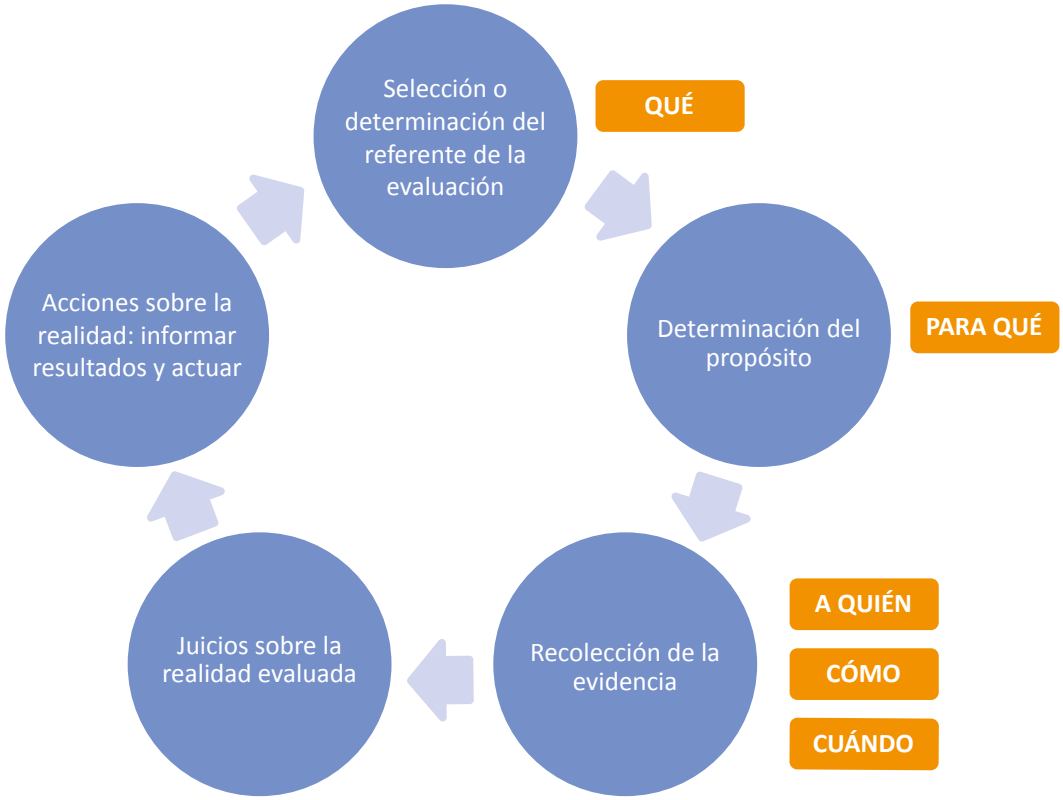
Desde el punto de vista de los estándares para la medición educativa y psicológica, la definición del dominio que será evaluado debe ser detallada y explícita desde un inicio, junto con una precisa declaración de las dimensiones de conocimiento, habilidades o procesos cognitivos, actitudes, valores o comportamientos que esperan abordarse. Esto, con el fin de resguardar la adecuación de las interpretaciones que se harán en el futuro de los resultados de la medición, en relación con los usos previstos de los mismos (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 2014).

Es tan crítica esta respuesta que, si el dominio a evaluar está pobremente definido o descuidadamente delimitado, ninguna acción que se desarrolle en las etapas posteriores del diseño de una evaluación, podrá subsanar su deficiencia (Downing, 2006).

En el ciclo de decisiones del diseño e implementación de una evaluación (ver figura 1), este cuadernillo referirá a la pregunta **¿Qué evaluar?**, que se vincula estrechamente con la pregunta **¿Para qué?**, y que es preciso responder previamente antes de definir el **cómo,** y el **cuándo.**

FIGURA 1

EL CICLO EVALUATIVO



Fuente: elaboración propia.

Entenderemos como **referente de la evaluación** al marco en el cual se circunscriben los modelos conceptuales o las declaraciones de contenido a los que se vincula el objeto de medida; es decir, es el marco de referencia del que se dispone para definir y delimitar los contenidos de la evaluación. Este marco puede existir con independencia de un determinado proyecto o programa evaluativo, y resultar más o menos explícito en relación al constructo evaluado. Lo distinguiremos del **marco de especificaciones**, en tanto este corresponde a una elaboración propia del diseño de la evaluación, y por lo tanto sirve a sus propósitos, delimitando y operacionalizando específicamente el contenido de la evaluación para guiar las definiciones metodológicas y la construcción de los instrumentos. El marco de especificaciones suele presentarse en un formato de tabla, donde se mapea en forma concreta el conjunto de contenidos, conocimientos y habilidades que serán abordados en el instrumento y el peso que se dará a cada uno de ellos a través de las preguntas (Ravela, 2006).

Se describirán en este cuadernillo estos componentes, fundamentando su relevancia y presentando distintos formatos que suelen encontrarse en los marcos de distintas evaluaciones. Proponemos un formato de tablas de especificaciones y sugerimos un conjunto de recomendaciones para elaborarlas.

I. Identificación del referente de la evaluación y definición del objeto de medida o constructo evaluado

Toda evaluación, cualquiera que sea su propósito, su magnitud y su contexto, conduce a la formulación de un juicio de valor, que establece la distancia entre la evidencia recogida y un estado o meta esperable en relación al constructo u objeto de medida[1]. Ya sea que evaluemos el aprendizaje de los estudiantes, el desempeño de los docentes, la gestión escolar en un centro educativo, o el clima laboral de una organización, en todos estos casos se establece un referente con el que comparamos los hallazgos de la evaluación y que contribuye a darles significado y establecer el juicio evaluativo. Usualmente, este referente tiene una connotación valorativa, pues contiene la descripción de un estado deseable en relación con la variable estudiada.

El marco de referencia que orienta una evaluación suele ser explícito; es decir, debe haber sido desarrollado de manera formal e institucionalizada con el fin de orientar los desempeños o los propósitos de mejora de un determinado fenómeno educativo. Es el caso de los marcos curriculares, que prescriben los aprendizajes esperados por los estudiantes en las distintas materias y niveles de enseñanza, o el de los estándares de desempeño docente, que han sido formulados para establecer explícitamente lo que debe saber y saber hacer un docente.

Un ejemplo de esto último es el Marco de Referencia para la Enseñanza (Framework for Teaching), desarrollado por Charlotte Danielson y colaboradores, a partir de su trabajo en el Educational Testing Service (ETS), que establece y describe un marco de competencias docentes, con el propósito de orientar tanto el desarrollo profesional como los procesos de evaluación del desempeño docente (Danielson, 2011). Este marco identifica cuatro dominios, reuniendo aquellos aspectos del ejercicio docente que inciden en el aprendizaje de los estudiantes (ver figura 2) y ha sido utilizado como base de sus sistemas de evaluación docente por muchos países latinoamericanos, o para inspirar la formulación de sus propios marcos de enseñanza, como es el caso de Chile y su Marco para la Buena Enseñanza (Ministerio de Educación [MINEDUC], 2008).

Estos marcos de referencia, si bien son precisos en definir y operacionalizar lo que se espera de una buena docencia, formulan sus definiciones de manera genérica, sin precisar los conocimientos y competencias particulares de los docentes de distintos niveles y materias, como sí ocurre con estándares específicos en Australia y algunos estados de la Unión Americana.

---

[1] Ver cuadernillo 1 de esta serie, sobre *Nociones básicas en medición y evaluación en el contexto educativo.*

FIGURA 2

DOMINIOS DEL MARCO PARA LA ENSEÑANZA

EL MARCO PARA LA ENSEÑANZA:

COMPONENTES DE LA PRÁCTICA PROFESIONAL

Dominio 1: Planificación y preparación	Dominio 2: El ambiente del aula
• Demostración de conocimiento del contenido y la pedagogía. • Demostración de conocimiento de los estudiantes. • Selección de los objetivos pedagógicos. • Diseño de una instrucción coherente. • Diseño de evaluaciones del aprendizaje de los estudiantes.	• Creación de un ambiente de respeto e interacción. • Establecimiento de una cultura para el aprendizaje. • Manejo de los procedimientos de aula. • Manejo de la conducta de los estudiantes. • Organización del espacio físico.
Dominio 3: Instrucción	**Dominio 4: Responsabilidades profesionales**
• Comunicación con los estudiantes. • Uso de técnicas de interrogación y debate. • Compromiso de los estudiantes en el aprendizaje. • Uso de evaluaciones en la instrucción. • Demostración de flexibilidad y sensibilidad.	• Reflexión sobre la enseñanza. • Mantenimiento de recursos rigurosos. • Comunicación con las familias. • Participación en una comunidad profesional. • Crecimiento y desarrollo profesional. • Muestras de profesionalismo.

Fuente: Danielson (2011).

Muchas evaluaciones, sin embargo, no cuentan con explicitaciones claras acerca de su ámbito de evaluación y deben utilizar como referentes modelos conceptuales o marcos de contenidos que fueron diseñados con otros fines, pero que resultan de utilidad para definir el objeto de medida y orientar las especificaciones de la evaluación. Como fue señalado, en el ámbito educativo surge la necesidad de medir una gran variedad de constructos o variables cuya definición está sujeta a varias perspectivas teóricas o modelos empíricos, y que demandan del evaluador la tarea de explicitar el referente tras una revisión y elaboración conceptual con tales fines. Ello es particularmente frecuente cuando se trata de medir variables psicosociales que no refieren a desempeños y que, por lo tanto, no han sido operacionalizadas con fines evaluativos; por ejemplo: actitud hacia el deporte, habilidades interpersonales, clima escolar, entre otras[2].

[2] Ver cuadernillo 5 de esta serie, *Desarrollo de instrumentos de evaluación: cuestionarios.*

La diferencia entre un marco de referencia implícito y uno explícito puede ser ejemplificada con dos posibilidades de evaluar la calidad de la gestión escolar a nivel de centros educativos. Por una parte, es posible utilizar para la evaluación de escuelas, un modelo conceptual a partir de los múltiples estudios de eficacia escolar que sustentan, teórica y empíricamente, cuáles son los factores que inciden en que los centros educativos obtengan buenos resultados a pesar de condiciones de contexto adversas. Murillo (2008) identifica y define ocho dimensiones a partir de esta revisión, entre ellas: (a) sentido de comunidad, que implica tanto la existencia de metas claras, conocidas y compartidas por parte de la comunidad escolar, como el trabajo colegiado por parte de los docentes; (b) liderazgo educativo, en que se destaca la importancia de un liderazgo más pedagógico que burocrático y se relaciona con las características de la persona que ejerce ese liderazgo y con su forma de ejercerlo; y (c) clima escolar y de aula, es decir, buenas relaciones entre los miembros de la comunidad educativa, y presencia de un ambiente afectivo y que fomente el trabajo de los alumnos en clase (Murillo, 2008, en Pérez, Ruiz, Langford y Prado, 2017). Estas definiciones pueden utilizarse como referentes, aunque requerirán un ejercicio de operacionalización y elaboración para guiar el diseño de la evaluación.

Comparativamente, la evaluación de centros educativos puede disponer de estándares o referentes explícitos que han sido desarrollados intencionadamente para la evaluación. Es el caso de la autoevaluación de escuelas en Hong Kong, en el que se cuenta con un marco de referencia explícito, generado por la Oficina de Educación, llamado Marco para el Desarrollo Escolar y Rendición de Cuentas (SDA, por sus siglas en inglés), y que contiene los criterios bajo los cuales se analizan internamente los centros educativos. Este marco de referencia comprende cuatro dominios, ocho áreas y un total de veintitrés indicadores de desempeño. Las áreas son las siguientes: administración escolar; liderazgo profesional; currículum y evaluación; aprendizaje y enseñanza; apoyo a estudiantes; colaboración; actitud y conducta; y participación y aprovechamiento académico (Pérez *et al*., 2017).

En el primer caso (modelos de eficacia escolar), el marco de referencia puede demandar al investigador una acuciosa revisión, análisis y síntesis de la literatura y estudios empíricos disponibles, la formulación de un modelo y, finalmente, a partir de lo anterior, la generación de especificaciones que orienten la construcción de los instrumentos. En el segundo caso, el SDA dispone ya de dominios y áreas identificadas, además de un conjunto de indicadores que describen operativamente lo que debería esperarse de las escuelas en los distintos ámbitos. La proximidad de este marco de referencia a los instrumentos es evidentemente mayor que en el caso anterior.

Si revisamos marcos de evaluaciones a gran escala de logro de estudiantes, también nos encontramos con casos que, por tener objetos de evaluación diversos, requieren echar

mano a marcos de referencia de distinta naturaleza. Uno de ellos es el Estudio Regional Comparativo y Explicativo, que lleva a cabo el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), con el propósito de evaluar la calidad de la educación, en términos de logros de aprendizaje, y factores asociados, en los países participantes de América Latina y el Caribe (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2016a). Para cumplir este propósito, el Tercer Estudio Regional Comparativo y Explicativo (TERCE) estudia el **desempeño escolar** general de los alumnos en escuela primaria, en las áreas de lenguaje, matemáticas y ciencias naturales, de los países participantes. Como estas variables en estudio corresponden a desempeño escolar, el referente de las pruebas se relaciona con el conjunto de aprendizajes que se espera que los alumnos logren en las asignaturas y niveles señalados, los cuales están explicitados en el currículum escolar. Como se trata de un estudio internacional, el marco de referencia son los currículum de los países participantes, en relación con los cuales se realiza un proceso de análisis que permite establecer los enfoques de enseñanza de las cuatro áreas evaluadas en cada uno, y la definición de ejes temáticos y procesos cognitivos que serán delimitados en las especificaciones de las pruebas (UNESCO, 2016b). El resultado es un instrumento que recoge evidencias del desempeño en tareas bastante semejantes a las acostumbradas a trabajar en la experiencia escolar de los evaluados.

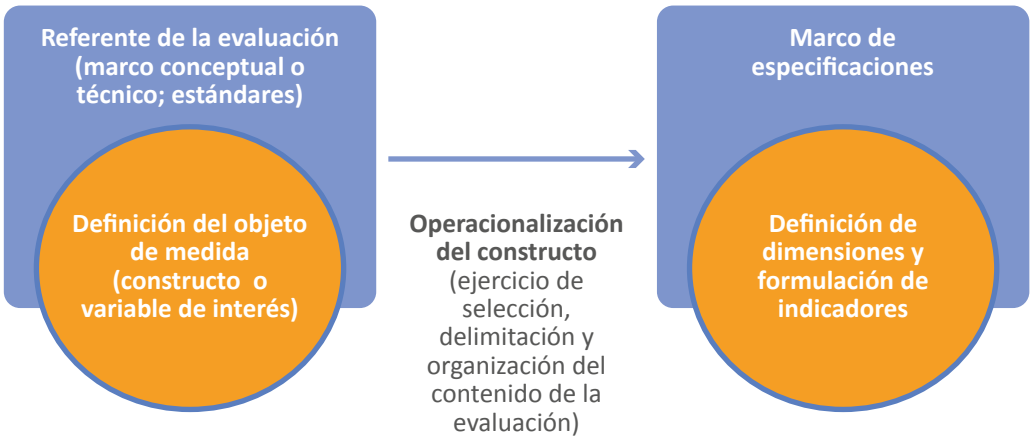
Distinto es el caso del Programa para la Evaluación Internacional de Alumnos (PISA) de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), cuyo objetivo es evaluar a los alumnos cuando llegan al final de la etapa de enseñanza obligatoria, hacia los quince años, en relación con las competencias requeridas para la vida; esto es: conocimientos y habilidades necesarios para la participación plena en la sociedad del siglo XIX (OCDE, 2016). Inicialmente se evaluaron tres ámbitos: lectura, matemáticas y ciencias, agregándose después la evaluación de competencias innovadoras, como resolución de problemas en contextos colaborativos y alfabetización financiera. El marco de referencia de PISA -dado su objeto de medida: las "competencias para la vida"-, no existe con independencia de esta evaluación, sino que se ha desarrollado para tales efectos con base en distintas fuentes. Además, debido a su carácter, ha ido actualizándose conforme el mundo evoluciona y las competencias para desenvolverse en él también cambian (OCDE, 2017). Este marco es elaborado con el concurso de un comité de expertos proveniente de los países participantes, que se dan a la tarea de definir y describir las competencias de los distintos ámbitos evaluados. Si bien la prueba considera las materias escolares básicas de ciencia, lectura y matemáticas, no las aborda en tareas en que los estudiantes deban reproducir el conocimiento, sino en aquellas en las que pueden extrapolar lo aprendido y aplicar ese conocimiento en circunstancias desconocidas, tanto dentro como fuera de la escuela y en contextos del mundo real. El enfoque de PISA considera "el hecho de que las economías modernas recompensan a los individuos no por lo que saben, sino por lo que pueden hacer con lo que saben" (OCDE, 2017, p.3).

Ya sea que se disponga de un marco de referencia preexistente y de alta pertinencia para una evaluación, o que este deba desarrollarse, se espera que a través de él pueda obtenerse una clara definición del constructo u objeto de medida. Insistimos en la importancia de este punto con respecto a evaluaciones de fenómenos complejos y multidimensionales, como la calidad del desempeño docente, (que queda claramente definida y descrita en el Marco para la Enseñanza de Danielson), o la calidad de la gestión escolar (que puede ser definida conceptualmente y operacionalizada mediante los modelos de eficacia escolar y los numerosos hallazgos empíricos que este campo de estudios ha generado).

Dejar claramente establecido el referente de la evaluación en las etapas iniciales del proceso conducirá a una precisa, detallada y comunicable definición del constructo u objeto de medida, y a orientar la siguiente tarea operativa: el desarrollo de las especificaciones de la evaluación, tal como se representa en la figura 3.

FIGURA 3

DEL MARCO DE REFERENCIA AL MARCO DE ESPECIFICACIONES



II. La operacionalización del constructo mediante el marco de especificaciones

Todavía en el desafío de responder a la pregunta del **qué** evaluar, y ya encaminándonos a la pregunta del **cómo** hacerlo, la tarea siguiente a la definición del constructo u objeto de medida es operacionalizar su significado al servicio de la evaluación. Esto significa tomar decisiones sobre el modo en que será entendido y observado en la situación evaluativa, y delimitar la extensión que se espera abarcar con la evaluación. Estas decisiones tienen como producto lo que se denomina **marco de especificaciones**, el cual funcionará como puente entre el referente y el instrumento o dispositivo con el que se levantará la información en la población o entidad evaluada.

Desarrollar y declarar las especificaciones antes de construir los instrumentos es una condición clave para garantizar que la medición logre una adecuada representación del constructo y que la información levantada permita efectivamente cumplir con los propósitos de la evaluación.

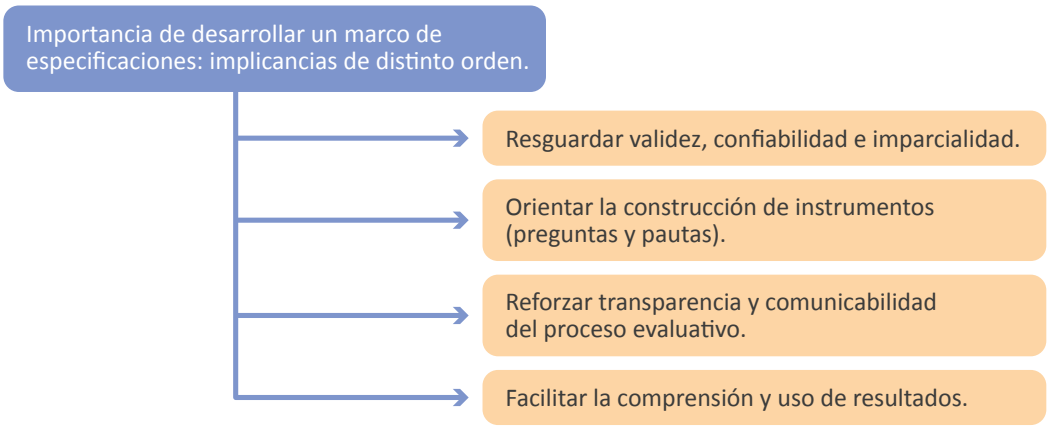
Aun cuando el marco de referencia sea muy explícito, la decisión de qué evaluar a través de un determinado instrumento sigue siendo una tarea decisiva. Sobre todo, en el caso de marcos complejos y multivariados, y más aún si se desea utilizar una batería de instrumentos.

Los riesgos de evaluar sin especificaciones de contenido

La importancia de desarrollar este marco de especificaciones se fundamenta en argumentos de distinta naturaleza, todos conducentes a resguardar la calidad técnica de la evaluación y su utilidad para los fines previstos. Algunos de estos se presentan en la figura 4.

FIGURA 4

IMPLICANCIAS DE DESARROLLAR LAS ESPECIFICACIONES DE LA EVALUACIÓN



En primer término, referimos al cumplimiento de los estándares técnicos, y a los resguardos de la validez, confiabilidad e imparcialidad que deben caracterizar toda medición (AERA, APA y NCME, 2014)[3]. Contar con un marco de especificaciones permite evitar los riesgos que más habitualmente atentan contra estos estándares: que los resultados no reflejen todos los aspectos relevantes del constructo de interés (subrepresentación), y que los resultados obtenidos estén explicados por aspectos no relacionados con el constructo de interés (varianza irrelevante). El primer riesgo es particularmente crítico cuando la evaluación se propone medir extensos dominios de contenido, que deben ser muestreados en diversos niveles de complejidad cognitiva, como es el caso de las pruebas de logro

[3] Más detalle en Cuadernillo 2 de esta serie, *Confiabilidad, validez e imparcialidad en evaluación educativa.*

de los estudiantes (cuyos marcos de referencia son extensos y multidimensionales). Elaborar especificaciones supone explicitar anticipadamente los aspectos (conocimientos, competencias, habilidades) del constructo que serán sometidos a evaluación. Tal representatividad del proceso de muestreo de contenido fortalece la evidencia de validez para las inferencias al dominio y reduce una de las principales amenazas a esta: representación escasa del constructo (Downing y Haladyna, 2006; Messick, 1989).

Se ha recogido evidencia abundante en la realidad educativa latinoamericana, que corrobora la presencia de esta debilidad a nivel de la evaluación de aula, administrada por los docentes: actividades de evaluación o instrumentos que se limitan a ofrecer tareas de complejidad inferior a las vinculadas con los aprendizajes que esperan evaluar, o dominios de contenidos que quedan excluidos (Loureiro, 2009; Ravela, Picaroni y Loureiro, 2017). Este hecho puede significar que la certificación de lo que los estudiantes saben y son capaces de hacer o no en un curso, no tenga relación con los contenidos y habilidades evaluados. La práctica de desarrollar especificaciones en la planificación de las pruebas permite visualizar si se están cubriendo todos los contenidos y habilidades pretendidos y reducir el riesgo descrito (Förster y Rojas-Barahona, 2008).

Otro riesgo para la calidad técnica por no contar con especificaciones es lo que Messick (1989) denomina "varianza irrelevante del constructo", que refiere al grado en que los puntajes de una prueba se ven afectados por procesos ajenos a lo que esta pretende medir (Downing y Haladyna, 2006). Esto ocurre cuando el desempeño del evaluado puede verse alterado porque la tarea evaluativa demanda poseer conocimientos o desplegar habilidades diferentes a las vinculadas al constructo; ello no solo resta validez a las posibles interpretaciones de los resultados, sino que puede ser fuente de inequidades para algún subgrupo evaluado y atentar, por tanto, contra la imparcialidad de la medición (AERA, APA y NCME, 2014). Por ejemplo, resolución de problemas matemáticos en contextos verbales extensos, que demandan habilidades de comprensión de lectura, además de las competencias matemáticas declaradas; o bien que refieren a temas más fácilmente abordados por hombres que por mujeres, como tablas de resultados futbolísticos, por ejemplo. Si bien estos problemas también se sitúan en otras etapas del desarrollo de los instrumentos (construcción y revisión de preguntas), igualmente pueden ser prevenidos con una adecuada definición del constructo y exhaustivas especificaciones de contenido.

En segundo término, una implicancia operativa de disponer de un marco de especificaciones que resulta de gran relevancia para quienes desarrollan instrumentos, es que la tarea de producción de preguntas o diseño de los reactivos de una actividad evaluativa se ve muy favorecida cuando se cuenta con orientaciones claras respecto al contenido de los mismos, tanto en términos de los dominios temáticos como de la complejidad de las

habilidades que las tareas que se solicitan al evaluado demandan. Por esto la gran mayoría de las evaluaciones a gran escala, que requieren una producción masiva de banco de ítems, disponen de marcos de especificaciones detallados y operativos (Argüelles, Castillo y Saragoni, 2015; Instituto Nacional de Evaluación Educativa [INEE], 2005).

Finalmente, la explicitación clara de los contenidos y habilidades evaluados es una condición imprescindible en el propósito de implementar evaluaciones transparentes y equitativas. Comunicar oportunamente a los evaluados el marco de especificaciones contribuye a clarificar los criterios que se encuentran a la base de la evaluación y a anticipar el tipo de tarea a la que se verán enfrentados. Esto es relevante cuando la evaluación tiene fines de certificación, pero también cuando el propósito es formativo, y se espera contribuir con información útil para retroalimentar a los protagonistas del proceso educativo.

Sobre este punto, Ravela y colaboradores (2017) alertan que la evaluación escolar diseñada y administrada por los docentes suele soslayar este requerimiento, enfocándose mucho más en los instrumentos y la evidencia, y escasamente en los referentes y criterios que la fundamentan: "en la evaluación el problema no es la subjetividad, sino la opacidad: no siempre explicitamos y comunicamos con claridad cuáles son los aspectos o dimensiones importantes, ni sobre la base de qué criterios y con qué evidencia estamos evaluando" (Ravela *et al*., 2017, p. 35). Esto conduce a evaluados que no comprenden con claridad el objeto de las evaluaciones y que difícilmente podrán aprovechar tal experiencia como una instancia informativa y formativa.

En estrecha vinculación con lo anterior, disponer de especificaciones de contenido precisas y comunicables favorece la comprensión y uso que puedan hacer los usuarios de la información levantada.

¿Cuáles son los componentes de un marco de especificaciones?

El contenido de las "especificaciones" que se encuentran documentadas en los marcos de las distintas evaluaciones, es muy variable. En algunos casos hace referencia a la finalidad, los usos previstos, los formatos de las preguntas y las características psicométricas que se esperan, y en otros se consideran hasta los modos de ejecución y administración de los instrumentos.

En este cuadernillo nos focalizaremos en las especificaciones que hacen referencia al contenido del instrumento y al modo como será distribuido este abordaje entre las distintas preguntas o tareas que lo componen.

Consideraremos, por tanto, los elementos que deben estar presentes en todo marco de especificaciones, y que a nuestro juicio le otorgan su carácter de orientador para la construcción y ensamblaje de los instrumentos. Estos elementos son:

- **Las áreas de contenido, dimensiones o dominios que constituyen el objeto de medida:** Ante constructos complejos o marcos de contenido muy extensos es necesario desagregar el dominio evaluado en partes que permitan organizar o secuenciar las especificaciones. Estos componentes suelen denominarse dimensiones, ejes o dominios y, dependiendo de la extensión del marco, pueden ser a su vez desagregados en subdimensiones, subejes o subdominios. Las dimensiones deben referir a aspectos distintos y relevantes del objeto de medida y ser, a la vez, complementarias entre sí.

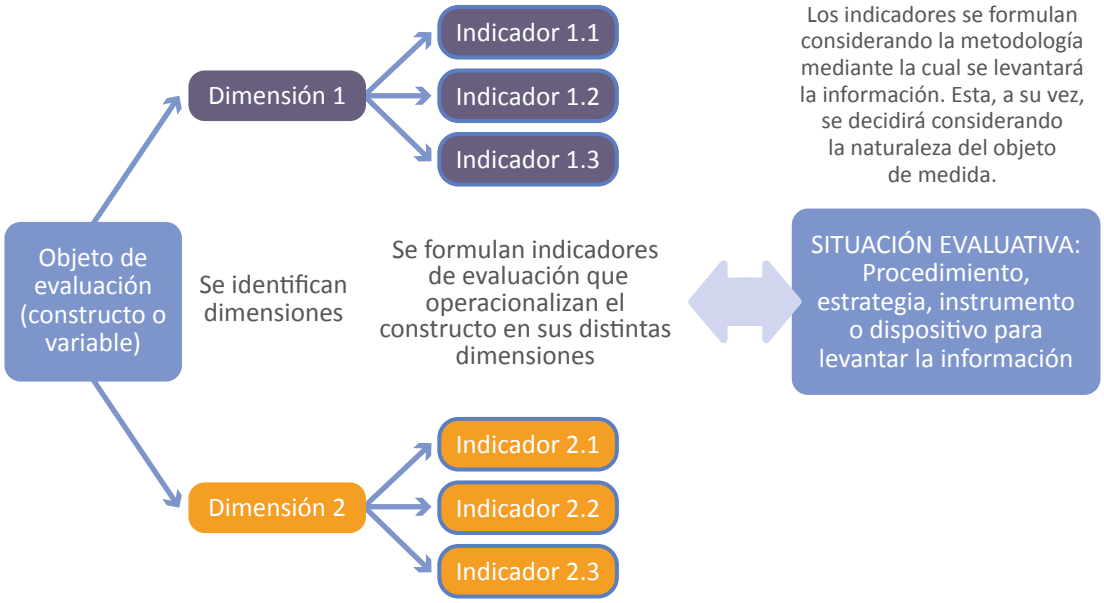
- **Los indicadores de evaluación:** Corresponden a la operacionalización concreta del constructo evaluado u objeto de medida, en tanto hacen referencia a tareas o desempeños observables que, en su conjunto, dan cuenta del atributo o del dominio que desea verificarse en la medición.

- **La planificación del instrumento:** Corresponde a las orientaciones relativas al diseño del instrumento y que guiarán su elaboración, considerando la ponderación que tendrá cada sección del marco y el número de preguntas asociadas a ellas.

La relación entre estos tres componentes se muestra en la figura 5.

FIGURA 5

RELACIÓN ENTRE LOS COMPONENTES DEL MARCO DE ESPECIFICACIONES



III. La tabla de especificaciones como herramienta que enmarca el desarrollo del instrumento

El contenido de las especificaciones para el desarrollo de un instrumento, suele ser vertido en una herramienta concreta que lo organiza seccionalmente en filas y columnas, y es usualmente llamada **tabla de especificaciones.** Ravela (2006) la define del siguiente modo:

> "La tabla de especificaciones es un instrumento para la elaboración de las pruebas. En él se consignan en forma esquemática los conocimientos, contenidos, objetivos, competencias (se emplean diferentes denominaciones...) que serán objeto de evaluación. Incluye además la indicación acerca de qué ítems o actividades de la prueba corresponden a cada contenido u objetivo... De esta manera, la Tabla de Especificaciones permite apreciar qué es lo que pretendía evaluar cada ítem de la prueba, permite garantizar que sean cubiertos todos los aspectos relevantes del referente, y permite apreciar el peso en cantidad de ítems que tiene cada aspecto". (Ravela, 2006, p. 113).

De esta definición se desprende que la tabla de especificaciones es una herramienta de mucha utilidad al momento del desarrollo de las preguntas y también cuando se seleccionan y organizan para el ensamblaje del instrumento.

Tanto la definición de los dominios como el desarrollo de posibles tareas que reflejan las especificaciones de la prueba, surgen del marco de referencia de la evaluación, y pueden utilizar como insumo información empírica, revisión de currículum, programas y planes de estudio o análisis conceptuales realizados por paneles de expertos (Webb, 2006).

La diversidad de tablas de especificaciones

La estructura que toma la tabla y su contenido pueden variar drásticamente de acuerdo con varios factores: el propósito de la evaluación y la naturaleza del objeto de medida, la especificidad del marco de referencia con que se cuenta, el dispositivo que se construirá e incluso el dominio o pericia de los desarrolladores de preguntas que deberán utilizar la tabla.

Respecto de lo primero, en el caso de instrumentos que se desarrollan para evaluar logros de aprendizaje, por ejemplo, tanto el contenido como los procesos cognitivos son aspectos importantes de las especificaciones de la prueba, y ambas características suelen reunirse en una tabla de especificaciones bidimensional, que cruzan ambas informaciones. De esta manera, se mapea el marco de referencia -los aprendizajes del currículo, digamos-, tanto en la cobertura temática, como en la diversa complejidad de habilidades que se proponen, con el fin de lograr una adecuada representatividad de aquél (Linn, 2006).

Así se presenta en el siguiente ejemplo, que corresponde a la tabla de especificaciones de la Prueba de Selección Universitaria de Chile de ciencias-biología. Esta evaluación cuenta con un extenso marco de especificaciones (el currículum de varios niveles de la educación secundaria), por lo que se delinean amplios dominios temáticos, cruzados por habilidades cognitivas, señalando la proporción de cada categoría que conformará el instrumento.

Ejemplo de tabla de especificaciones: Prueba de Selección Universitaria de Chile (Ciencias-Biología, módulo común)

Área temática	Habilidad Cognitiva				%
	Reconocimiento	Comprensión	Aplicación	Análisis, síntesis y evaluación	
Organización, estructura y actividad celular					30%
Procesos y funciones vitales					13%
Biología humana y salud					18%
Herencia y evolución					16%
Organismo y ambiente					23%
TOTAL	40%		60%		100%

Fuente: recuperada de http://www.psu.demre.cl/la-prueba/pruebas-y-temarios/temario-prueba-ciencias-tabla-especificaciones.php

En el siguiente ejemplo, se presenta la tabla de un instrumento que aborda un contenido más específico, Lectura, de un solo nivel, 6° grado, por lo que identifica áreas temáticas y habilidades de mayor especificidad.

Ejemplo de tabla de especificaciones: Tercer Estudio Regional Comparativo y Explicativo, TERCE (Lectura, 6° grado)

Dominio	Proceso			Total	%
	Literal	Inferencial	Crítico		
Comprensión intratextual	23	48	1	72	75%
Comprensión intertextual	3	5	3	11	11%
Metalingüístico y teórico	0	13	0	13	14%
TOTAL	26	66	4	96	100%
%	27%	69%	4%	100%	...

Fuente: UNESCO (2016 a).

En los dos casos anteriores se cuenta con marcos de referencia bastante explícitos, que detallan con precisión los aprendizajes esperados y sus correspondientes contenidos temáticos y habilidades. Estos marcos se utilizan complementariamente a la tabla como guía para la elaboración de las preguntas.

Sin embargo, cuando no se cuenta con un marco de referencia que pueda utilizarse directamente, los desarrolladores del instrumento deben hacerse cargo de un ámbito mayor de decisiones en la especificación del contenido; por ejemplo, cómo acotar y organizar las áreas temáticas y cómo será operacionalizado el constructo en cada una de ellas. En dichos casos, resulta de utilidad contar con una declaración más específica de lo que desea evaluarse y del modo en que se recogerá la evidencia de desempeño en las distintas tareas que proponga el instrumento a través de sus preguntas. Los indicadores de evaluación cumplen este papel, integrando los contenidos y habilidades, y entregando una mejor orientación tanto para clarificar los criterios de evaluación, como para guiar la tarea de los elaboradores de preguntas. El siguiente ejemplo muestra una tabla con estas características, destinada a evaluar a estudiantes egresados de un programa pedagógico de matemáticas. Se presenta solo una porción de ella, pues, como es de suponer, en este caso se trata de tablas de una mucho mayor extensión que las anteriores.

Ejemplo de porción de tabla de especificaciones para egresados de Pedagogía en Matemáticas

Dominio	Subdominio	Indicadores	Ponderación
Números y álgebra	Razones	Explican el concepto de razón y la información que las razones entrega.	20%
		Interpretan en diversas situaciones el concepto de razón.	
		Explican una razón a través de una representación concreta o una pictórica.	
		Reconocen las variables presentes en una razón.	
	Proporciones	Comparan cocientes de dos razones para formar una proporción.	20%
		Usan regla de tres para conocer el término desconocido de una proporción.	
		Establecen proporciones y las usan para resolver diversas situaciones de variación proporcional.	
	Variación proporcional directa	Reconocen situaciones cotidianas donde se presentan proporciones directas.	30%
		Determinan la constante de proporcionalidad y la utilizan para realizar cálculos.	
		Representan en tablas y gráficos variaciones proporcionales directas.	
	Variación proporcional Inversa	Reconocen situaciones cotidianas donde se presentan proporciones inversas.	30%
		Determinan la constante de proporcionalidad y la utilizan para realizar cálculos.	
		Representan en tablas y gráficos variaciones proporcionales inversas.	

Fuente: elaboración propia, a partir de la evaluación de egresados de Pedagogía de una universidad latinoamericana.

Otros formatos y contenidos se pueden encontrar en las tablas de especificaciones utilizadas con el fin de construir instrumentos distintos, como los cuestionarios para evaluar opiniones y actitudes mediante encuestas. En estos casos, no se trata de medir aprendizajes o rendimiento, por lo que las especificaciones no refieren a habilidades o desempeños esperados, sino a los contenidos abordados y la proporción en la que las preguntas del instrumento los cubren. En el ejemplo a continuación se aprecia que la tabla incluye también la escala con la que será juzgado cada uno de los aspectos.

Ejemplo de tabla de especificaciones de una encuesta de satisfacción en los estudiantes del primer año de la carrera de médico cirujano, UNAM, México

Número	Aspecto a evaluar (categorías)	Ítem	Escala
1	Satisfacción general en el primer año de la carrera.	1	Muy satisfactorio, satisfactorio, insatisfactorio, muy insatisfactorio.
2	Tiempo en horas (por asignatura).	1	Muy adecuadas, adecuadas, excesivas, escasas.
3, 4, 5	Aspectos didácticos (por asignatura). Integración básico-clínica. Claridad de la bibliografía sugerida. Uso de recursos y materiales que facilitan el aprendizaje.	3	Casi siempre, muy a menudo, de vez en cuando, casi nunca.
6	Si la evaluación refleja el aprendizaje (por asignatura).	1	Casi siempre, muy a menudo, de vez en cuando, casi nunca.
7	Si considera que los contenidos le serán de utilidad (por asignatura).	1	Casi todos, la mayoría, solo algunos, casi ninguno.
8	Satisfacción (por asignatura).	2	0-25%, 26-50%, 51-75%, 76-100%.
	Datos de identificación.	2	Edad y sexo (F, M).

Fuente: Osorio-Alvarez y Parra (2015).

En este tipo de estudios también es posible agregar otros campos que contribuyen a especificar el diseño de los instrumentos y el análisis que se hará de los resultados, como enunciación de la evidencia mediante la que se recogerá información sobre cada tema (que actúa en este caso como indicador), o el tipo de variable al que corresponde cada atributo medido[4].

[4] Mayor detalle en cuadernillo 5 de esta serie, Desarrollo de instrumentos de evaluación: cuestionarios.

Por ejemplo, si se desea estudiar mediante una encuesta la accesibilidad de los servicios de salud de los habitantes de una localidad, se puede considerar como un indicador de la accesibilidad geográfica el tiempo que toma a las personas llegar desde sus hogares a los centros de salud. Ese tipo de información será explorada mediante las preguntas del cuestionario. Ver el siguiente ejemplo.

Ejemplo de tabla de especificaciones de un cuestionario para evaluar la accesibilidad a los servicios de salud de una localidad

Constructo o variable en estudio	Definición del constructo	Dimensiones	Indicadores	Número de preguntas	Ejemplos de preguntas
Accesibilidad de los servicios de salud	Oportunidades y restricciones de los habitantes de una localidad para tomar contacto con los centros de salud disponibles y aprovechar los servicios que estos pueden ofrecer.	Accesibilidad geográfica	• Tiempo de traslado desde el hogar al centro de salud. • Medios de transporte disponibles entre el hogar y el centro de salud.	10	¿Por qué medio de transporte llega al centro de salud? ¿Cuánto tiempo tarda en llegar desde su hogar al centro de salud?
		Accesibilidad económica	• Montos de dinero requeridos para la atención. • Disponibilidad de dinero para atención de salud.	15	¿Requiere hacer pagos para ser atendido en el centro de salud?
		Accesibilidad cultural	• Conocimiento sobre el tipo de atención que ofrece el centro de salud. • Conductas de búsqueda de información sobre la atención de salud.	25	¿En qué situaciones ha debido acudir al centro de salud? ¿Qué otras atenciones sabe usted que ofrece el centro de salud?

Fuente: elaboración propia.

Finalmente, otro factor que incide en la información que se entrega en un marco de especificaciones, tiene que ver con la necesidad de guiar más o menos directivamente la construcción de las preguntas. En los casos en que se requieren varias formas paralelas de los instrumentos, como ocurre con las evaluaciones estandarizadas a gran escala, además de las herramientas estadísticas que permiten garantizar la comparabilidad, es recomendable contar con diversas versiones equivalentes de ítems o preguntas. Esto se favorece desarrollando matrices que especifiquen, aún con mayor detalle que en los ejemplos dados, el contenido de las preguntas y las tareas que solicitan. Se describe, entonces, el tipo de contexto, situación problema, su contenido, extensión y complejidad, sobre los que deben versar las preguntas. Ello permite uniformar las propuestas de los elaboradores de preguntas y reducir los efectos individuales en el producto de su trabajo.

La ventaja de contar con mayores especificaciones para la elaboración de los ítems, tiene a su vez una desventaja equivalente: la homogeneidad de las preguntas. Como señala Webb (2006), la redacción de ítems o reactivos tiene un componente de arte, que está dado por la creatividad de sus autores; mientras mayores sean las especificaciones, mayor será la estandarización y menor la diversidad de preguntas disponibles para conformar el instrumento.

Identificar el contenido de una prueba diseñada para medir el conocimiento y las habilidades de los estudiantes, es tanto un arte como una ciencia. La ciencia de la especificación de contenidos se basa en marcos conceptuales, modelos matemáticos y procedimientos replicables. El arte de la especificación de contenidos se basa en juicios de expertos, en la redacción de ítems de prueba efectivos y en el equilibrio de los muchos elementos que deben compensarse (Webb, 2006).

Una guía de pasos para elaborar una tabla de especificaciones de una prueba

A continuación, presentamos un modelo de trabajo que, a juicio de las autoras de este cuadernillo, y con base en la experiencia, resulta de utilidad para el propósito de elaborar las especificaciones de una prueba. Se entregan y fundamentan las recomendaciones para cada uno de los pasos y se distinguen los resguardos que deben cuidarse en cada uno de ellos. En la sección siguiente se ofrece una lista de cotejo que permite verificar el cumplimiento de los criterios señalados, a modo de herramienta para evaluar la calidad y operatividad de la tabla construida.

Consideraciones generales

Antes de detallar los pasos de este proceso, es necesario declarar los fundamentos generales que sustentan algunas de las recomendaciones. Estas bases se configuran como tales no solo a partir de la evidencia conceptual y empírica disponible en la literatura, sino también de los aprendizajes logrados a través de una extensa experiencia en construcción de instrumentos de evaluación, poblada de aciertos y desaciertos, de preguntas y respuestas.

- Recomendamos una tabla integrada, en la que el **contenido** no sea disociado de las **habilidades cognitivas** que se ponen en juego para demostrar su dominio, ni se mapeen en forma separada. Particularmente en el caso de pruebas que evalúan desempeños educativos, referidos a aprendizajes que contienen de manera integrada ambos componentes, existen mejores probabilidades de obtener una muestra representativa en la medida en que se especifiquen directamente posibles tareas que los operacionalizan. Este papel lo cumple el conjunto de indicadores de evaluación que, debido a ello, consideramos un elemento nuclear en una tabla de especificaciones.

  Las habilidades no pueden desplegarse en ausencia de contenidos, y en la mayoría de los casos no es de interés de los sistemas educativos, ni de los docentes en el aula, promoverlas –y, por lo tanto, evaluarlas– de forma aislada (acaso este sea un objeto de estudio de la investigación en psicología cognitiva, pero no del proceso educativo). Asegurar la cobertura del amplio abanico de aprendizajes, en cuanto a complejidad y profundidad, es más efectivo si se formulan indicadores que respondan a esa progresión en el marco de especificaciones de un instrumento, que asociar ítems a las celdas de una determinada taxonomía de habilidades.

- El desarrollo de una tabla de especificaciones es una ardua tarea de selección, priorización y organización de contenidos, y requiere ser llevada a cabo por equipos especialistas conocedores de la materia evaluada, del marco de referencia y del contexto donde la evaluación se sitúa. Sin embargo, también es fundamental que el trabajo sea guiado por expertos en la construcción de instrumentos de evaluación. Como se describirá enseguida, la formulación de indicadores debe tener a la vista los alcances y limitaciones del tipo de instrumento que se está construyendo y el formato de las preguntas que lo componen. No es lo mismo formular indicadores si se evaluará el constructo mediante una prueba escrita de ítems de opción múltiple, que si se evaluará mediante una pauta de observación para apreciar directamente el desempeño de los estudiantes. Estas distinciones suelen ser invisibles para expertos en el contenido, que no tienen experiencia en el diseño de instrumentos. Asimismo, no es igual desarrollar un marco de especificaciones con un propósito de diseño curricular (para elaborar mallas o programas de un curso, por ejemplo), desafío que

un experto disciplinario podría enfrentar exitosamente por sí mismo, que un marco de especificaciones al servicio del diseño de un dispositivo para levantar información en el contexto de una evaluación.

- Por otra parte, el levantamiento de un marco de especificaciones supone, además de un conjunto de decisiones de orden disciplinario y técnico, decisiones políticas, sobre todo si se trata de evaluaciones de altas consecuencias. Definir qué extensión y profundidad del contenido debe evaluarse, y en qué ponderación establecer su presencia en el instrumento, puede tener implicancias políticas. Por todo lo anterior, sostenemos que, en el proceso que describiremos, debería contarse con instancias de participación, revisión y validación de actores de los distintos perfiles: especialistas en el contenido, en el diseño de instrumentos, y tomadores de decisiones.

- Como toda herramienta del campo de la evaluación en particular, y de la educación en general, las tablas de especificaciones solo pondrán en juego su calidad cuando sean utilizadas; esto es, cuando hayan cumplido el papel de orientar un proceso de construcción de preguntas y desarrollo de un instrumento. En su uso quedan en evidencia sus debilidades o problemas: la selección de su contenido, la claridad y adecuación de sus indicadores y el grado en que facilitaron o dificultaron la generación de preguntas, su equilibrio en términos de la especificidad lograda en las distintas áreas temáticas, etcétera. Por ello sostenemos que las tablas de especificaciones, aunque pueden usarse como un elemento estable en el tiempo para evaluaciones periódicas, deben estar en permanente revisión, internalizando la retroalimentación que entrega el proceso de construcción de preguntas que guía, los resultados obtenidos de la evaluación y, en caso de existir, el proceso de establecimiento de puntos de corte. Y por supuesto, recogiendo las modificaciones que el marco de referencia de la evaluación pudiese sufrir en el tiempo.

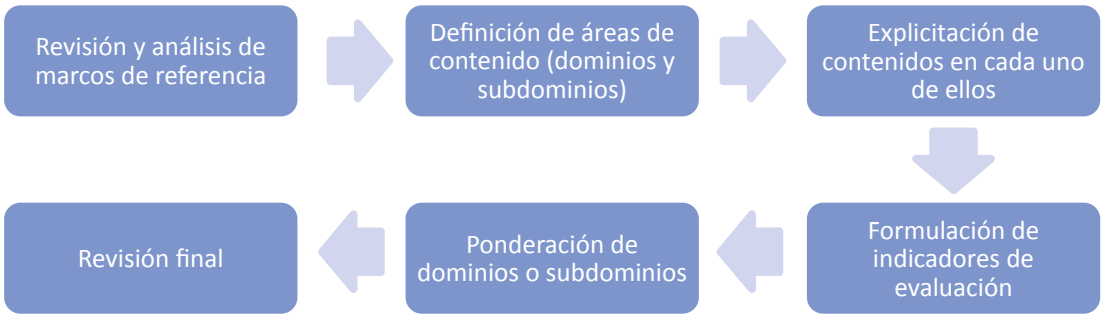
- Finalmente, declaramos explícitamente nuestra predilección por la simplicidad sobre la complejidad. La tabla de especificaciones de un instrumento debe ser una herramienta de trabajo amigable y útil para los elaboradores de preguntas, y clara e informativa para los demás usuarios de la evaluación y sus resultados. Excesivo cruce de atributos o multiplicidad de niveles o variables, con el afán de representar la complejidad de un constructo, pueden atentar contra la usabilidad de esta herramienta. Ello no quita que pueda contarse con documentos complementarios en los cuales se desarrollen mayores explicaciones o descripciones de su contenido, pero la tabla en sí debería ofrecernos una mirada sinóptica que clarifique el abordaje del objeto de medida a través del instrumento.

Los pasos en la elaboración de una tabla de especificaciones

El proceso de desarrollo de una tabla de especificaciones puede sintetizarse en los siguientes pasos que se representan en la figura 6, algunos de los cuales ya han sido comentados en las páginas anteriores, y se describen a continuación.

FIGURA 6

PASOS PARA LA ELABORACIÓN DE UNA TABLA DE ESPECIFICACIONES



1. La revisión y análisis de los marcos de referencia son particularmente relevantes cuando la evaluación aborda un constructo novedoso o escasamente medido. Como ya se señaló, si se cuenta con un marco de referencia explícito que distinga y defina los objetivos de evaluación, un cuerpo de estándares, marcos de contenido o desempeños, esta etapa se limitará a la revisión y profundización de tales herramientas. Si, en cambio, el marco debe elaborarse y explicitarse, la tarea de revisión se extiende a todos los documentos relacionados con el objeto de medida que puedan contribuir a definirlo y establecer su contenido para especificar su evaluación. Por ejemplo, la elaboración de pruebas destinadas a evaluar el conocimiento que poseen los profesores sobre la materia que enseñan, en la mayoría de los países que las aplican como parte de sus políticas docentes, no cuentan con referentes de estándares específicos sobre este constructo. Es decir, no existe una declaración central explícita acerca de lo que debería saber y saber hacer un docente de determinada asignatura para lograr con sus estudiantes los aprendizajes del currículum (Martínez, 2016). Esto exige que las especificaciones de los instrumentos sean antecedidas por un exhaustivo proceso de revisión de referentes relevantes disponibles: el primero de ellos, el currículum escolar, que explicita los aprendizajes que los docentes deberían promover en sus estudiantes –mas no lo que deberían saber para lograrlo-; también las orientaciones didácticas que en documentos oficiales se hayan puesto a disposición de las escuelas y maestros por parte de las autoridades educativas; documentos o estándares que orientan la formación inicial de los maestros, etcétera. Es decir, a partir de diversas fuentes, se reúnen los insumos para la determinación de los contenidos y ámbitos que serán evaluados y su posterior organización en la tabla de especificaciones.

2. Definición de áreas de contenido (dominios o dimensiones): como ya fue señalado, en los casos en que el objeto de medida es extenso o multidimensional, es recomendable que el contenido sea organizado en las áreas o subáreas que puedan desagregarse y ser representadas en la tabla de especificaciones. Usualmente, también será útil distinguir subáreas de contenido. Sugerimos denominar a estas desagregaciones **dominio y subdominios temáticos** cuando se trata de evaluar un marco de conocimientos, y **dimensión o subdimensión** cuando se trata de evaluar constructos psicosociales más abstractos.

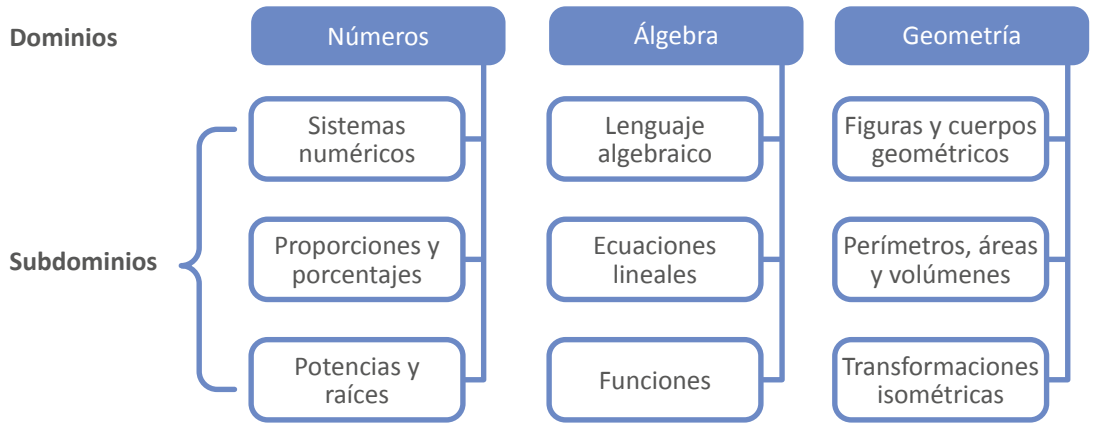
La tarea de desagregación de las áreas de contenido es sencilla cuando existen marcos de referencia explícitos o marcos conceptuales validados, que suelen sugerir esta subdivisión, mas requiere un ejercicio de análisis y síntesis importante en caso de no contar con ellos.

Los currículum educativos suelen dividir las materias en áreas o ejes temáticos (como se ejemplifica en la figura 7), y los marcos de especificaciones de las pruebas que evalúan aprendizajes escolares suelen alinearse a dichas categorías.

Cabe señalar que la necesidad de desarrollar este paso está estrictamente vinculada con la amplitud del marco de contenidos, por lo que, en evaluaciones parciales de unidades acotadas de aprendizajes, por ejemplo, podría ser innecesario desagregar subdominios e incluso dominios, desprendiéndose los indicadores directamente desde el objeto de evaluación.

FIGURA 7

ÁREAS TEMÁTICAS DEL CURRÍCULUM DE MATEMÁTICAS DEL SEGUNDO CICLO DE LA EDUCACIÓN BÁSICA, CHILE



En el caso de evaluaciones de constructos abstractos o teóricos, las desagregaciones deberán basarse en los marcos conceptuales revisados, como el ejemplo dado en las secciones anteriores para medir la gestión escolar con base en modelos de eficacia escolar. Si el objeto de medida ha sido escasamente explorado, y es necesario levantar desde cero el marco de especificaciones, las áreas de contenido deberán ser propuestas como un ejercicio innovador fundado en la revisión de los referentes, la información empírica disponible y la participación de expertos en el contenido.

En este caso, las tablas de especificaciones pueden elaborarse mediante un proceso inductivo en el que, a partir de la formulación de los indicadores de evaluación, se identifican las diferencias entre ellos y se agrupan por la consistencia de contenido que los une, conformando las dimensiones o subdimensiones. O bien, a partir de las categorías ya distinguidas (a partir de los marcos de referencia), se formulan los indicadores para cada una de ellas. En el siguiente recuadro se muestran ejemplos de estos caminos.

ANALIZANDO Y ESPECIFICANDO EL CONSTRUCTO EVALUADO:
El levantamiento de dimensiones o dominios

Se pueden tomar dos caminos con el propósito de especificar áreas de contenido de un constructo muy extenso o complejo, que organicen el conjunto de indicadores que lo operacionalizarán. Uno de ellos consiste en utilizar una **estrategia deductiva**, extrayendo las dimensiones o dominios del marco de referencia que las sugiere y desprendiendo los indicadores a partir del contenido que aquél declara. El segundo camino, en ausencia de un marco de referencia disponible, es utilizar una **estrategia inductiva**, es decir, formular indicadores que refieran al constructo y a partir de su análisis identificar los de contenido común y agruparlos para conformar las dimensiones.

Ejemplo de estrategia deductiva: Si se desea construir un instrumento para una evaluación sumativa de un curso de Literatura Universal, es posible utilizar como dominios las categorías que organizan el curso y que suelen utilizarse en este ámbito disciplinario: Edad Antigua, Edad Media, Renacimiento, Edad Moderna, Contemporánea. Con base en cada una de ellas, se formularán los indicadores.

Ejemplo de estrategia inductiva: Se desea construir un examen de conocimientos y competencias para ser utilizado en la selección de postulantes a un cargo de bibliotecario de una determinada institución educativa, sin embargo, no se cuenta con la descripción de cargo y no hay experiencias previas de esta evaluación. Para construir el marco de especificaciones se proponen indicadores de un buen bibliotecario, en términos del conocimiento y habilidades que debe poseer (puede ser también recogida de los usuarios internos y externos de la biblioteca). A partir del contenido de estos indicadores preliminares, se identifican áreas comunes que conformarán los dominios o dimensiones y, eventualmente, subdominios o subdimensiones si se identifican categorías de distinto nivel lógico o extensión.

3. Explicitación de contenidos de dominios o dimensiones: sugerimos listar los contenidos que esperan abordarse en cada dominio o subdominio, pues esto facilita la tarea de monitorear la cobertura temática de la tabla en la formulación de indicadores. Como ya fue señalado, en el caso de pruebas con referentes extensos, que deben mapear contenidos de varias unidades o niveles educativos, el desafío es llevar a cabo un buen ejercicio de selección y priorización, pues la decisión de qué se incluirá y qué se excluirá de la evaluación tendrá consecuencias para los evaluados y, por tanto, debe estar debidamente fundamentada (Ravela *et al*., 2017).

En términos prácticos, para conservar la legibilidad de la tabla, si este listado es muy extenso, puede ir dispuesto en un documento complementario y no al interior de esta.

4. Formulación de indicadores de evaluación: los indicadores son la descripción cualitativa o cuantitativa concreta (observable) de la presencia de determinado atributo o del logro de un determinado objetivo. Como ya fue declarado, el levantamiento de los indicadores es la tarea central del proceso de elaboración de tablas de especificaciones, pues estos actúan como el eslabón entre el objeto de medida y el instrumento mediante el cual se recogerá la información sobre él. Los indicadores echan luz sobre los desempeños, acciones o eventos que serán considerados como evidencia del constructo evaluado en cada una de las dimensiones o dominios.

En las pruebas que evalúan aprendizajes, los indicadores se formulan como expectativas de logros y representan acotados desempeños que se constituyen en muestras de dichos aprendizajes.

En los indicadores se integran elementos de contenido y habilidades, en tanto se enmarcan dentro de las temáticas abordadas por el dominio o subdominio al que pertenecen, pero se explicita el modo en que se espera que el evaluado opere con ellos. De esa forma, también hacen referencia a procesos cognitivos involucrados en la tarea sugerida por el indicador, pero siempre puesto en ejecución en una acción observable, apreciable en una situación evaluativa.

Los atributos de un **indicador correctamente formulado** son los siguientes cuatro:

- Refiere a un **desempeño o condición observable** directamente a través de la evaluación. En tal sentido no recomendamos indicadores que aludan a procesos mentales o a conceptos abstractos. El valor del indicador es operacionalizar el objeto de evaluación y eso significa, en palabras sencillas, hacer concreto lo abstracto. En muchos casos

se desea evaluar aprendizajes o variables relacionados con procesos mentales de los evaluados, justamente en esos casos el papel del indicador es traducirlos en conductas o acciones observables (ver tabla 1).

- Contiene un **referente único, está formulado de modo claro y, en lo posible, en forma positiva** (lo que se espera que el evaluado sepa, no señalando aquello que no sabe). Se deben evitar complejidades innecesarias en el uso del lenguaje, de manera que cualquier usuario de la tabla interprete inequívocamente la tarea a que refiere (ver figura 8).

- Guarda una relación de **correspondencia con el dominio y subdominio** al que representa y, en consecuencia, con el objetivo de la evaluación.

- El conjunto de indicadores **debe dar cuenta, en extensión y profundidad, del objeto evaluado**.

Cumplir con las condiciones descritas permite que los indicadores se constituyan en evidencia de validez de contenido de los instrumentos y contribuyan a minimizar los riesgos descritos en la primera sección de este cuadernillo: que el constructo no quede adecuadamente cubierto y que los resultados de la evaluación no den cuenta efectivamente de lo que esta declara medir y no de otras variables interferentes.

5. Ponderación de dominios y subdominios: Finalmente, se debe establecer la proporción del instrumento que será destinada a cada área de contenido identificada; esto es, el porcentaje de preguntas mediante el cual se cubrirá cada dominio o subdominio de la tabla. En tablas pequeñas o acotadas, esta ponderación podría asignarse directamente a los indicadores. La decisión de la cobertura debe estar basada en criterios técnicos o disciplinarios, pero debidamente fundamentados, sea por la extensión de los contenidos, por la relevancia asignada en el marco de especificaciones o por el juicio lógico de los expertos. Como todo en el ciclo evaluativo, es una decisión que depende del contexto y que no tiene una única respuesta, pero siempre debe conservar como foco la adecuada cobertura y representación del objeto de medida.

TABLA 1

ERRORES EN LA FORMULACIÓN DE INDICADORES

Errores posibles en los indicadores	Indicador con error	Indicador bien formulado
No plantearlos en términos observables: Los indicadores deben plantearse como tareas verificables mediante la evaluación.	Comprende el sentido global del texto.	Expresa con sus palabras el sentido global del texto.
	Valora el cuidado de su entorno próximo.	Reconoce los beneficios de cuidar su entorno próximo.
	Empatiza con las necesidades de sus pares.	Escucha activamente cuando sus pares expresan sus necesidades.
Formularlos con más de un referente: Es conveniente plantearlos con un solo referente. Cuando refieren a dos asuntos distintos, es mejor plantear dos indicadores distintos.	Describe el funcionamiento del sistema digestivo y reconoce las partes principales de los órganos involucrados.	Indicador 1: Describe el funcionamiento del sistema digestivo. Indicador 2: Identifica las partes principales de los órganos que componen el sistema digestivo.
Usar un lenguaje o una formulación innecesariamente compleja: Es necesario que los indicadores sean claros, directos y, en lo posible, formulados en positivo.	Enfatiza aspectos motivacionales del trabajo colectivo, estableciendo prioridades del proceso riguroso por sobre el producto.	Anima activamente a su grupo para llevar a cabo un proceso de trabajo riguroso.
	Utiliza oraciones desprovistas de sentido.	Formula oraciones claras y coherentes.

Criterios para verificar la calidad de una tabla de especificaciones

La calidad de una tabla de especificaciones debe ser verificada en función del rol que cumple; es decir, del grado en que proporcionan una guía adecuada para el desarrollo de los elementos del instrumento. Aunque en su construcción hayan participado confiables expertos, esto no es garantía de que las especificaciones que propone sean las adecuadas. Es por ello, que requieren ser revisadas y evaluadas (Downing, 2006).

En la tabla 2 se presenta un conjunto de preguntas guía para la revisión final de una tabla de especificaciones; esta pauta puede ser aplicada por los mismos desarrolladores del instrumento, pero es recomendable que también lo hagan jueces externos al equipo.

TABLA 2

PAUTA PARA LA REVISIÓN FINAL DE UNA TABLA DE ESPECIFICACIONES

Sobre las dimensiones o dominios	• ¿Representan los dominios y subdominios la totalidad del constructo u objeto de medida? ¿Se excluyen áreas de contenido relevante de acuerdo con la definición del constructo y el marco de referencia? • ¿Los dominios y subdominios son exhaustivos (no se traslapa su contenido)? • ¿El nombre de los dominios y subdominios es el apropiado a su contenido?
Sobre los indicadores	• ¿Son los indicadores coherentes con el dominio y subdominio al cual pertenecen? • ¿Son los indicadores un indicio claro del cumplimiento o presencia del objeto de medida? • ¿Refieren los indicadores a conductas observables, y por lo tanto medibles, mediante preguntas del formato definido? • ¿Son los indicadores necesarios y suficientes para abordar a cabalidad y de modo relevante la totalidad del constructo? • ¿Son claros los indicadores y pueden ser comprendidos inequívocamente por distintos usuarios? • ¿Evocan un solo desempeño o cumplirlos involucra más de una capacidad en el evaluado? • ¿Son muy acotados los indicadores, de manera que solo se desprende de ellos una sola tarea o pregunta?
Sobre las ponderaciones	• Las ponderaciones establecidas a dominios y subdominios, ¿corresponden al referente de la evaluación? • Si responden a otro argumento, ¿se fundamenta en la mejor representatividad del constructo?
Sobre la tabla en su globalidad	• En su conjunto, ¿la tabla de especificaciones da cuenta del objetivo de evaluación o del constructo evaluado? • ¿Está adecuadamente representada la complejidad del constructo? • ¿Tiene una extensión equivalente a la extensión o complejidad del constructo? • ¿Tiene una extensión proporcional al instrumento(s) de evaluación? • ¿Representa una guía clara y útil para la construcción de las preguntas?

Consideraciones finales: ideas fuerza

- Se ha fundamentado en este cuadernillo la importancia de explicitar el marco de referencia que orienta un determinado proceso evaluativo y la necesidad de definir y operacionalizar con precisión el objeto de medida o constructo que se desea conocer. Tal como lo establecen los estándares para pruebas educativas y psicológicas, la especificación del contenido de las evaluaciones mediante estas acciones es una condición que permite resguardar la adecuación de las interpretaciones que se harán de los resultados de la medición. Por su parte, al omitirlas, se arriesga la infrarrepresentación del constructo o la ocurrencia de varianza irrelevante en dichos resultados, debido a la interferencia de factores indeseados en el desempeño de los evaluados.

- Se distingue la necesidad de contar con un marco de referencia de la evaluación, del cual se desprendan los modelos conceptuales o el contenido declarativo implicado en el objeto de medida; y a partir de este, desarrollar un marco de especificaciones, al servicio de la elaboración de los instrumentos, que permita delimitar y operacionalizar los contenidos y habilidades que esperan evaluarse.

- Se reconoce una gran diversidad de formatos y contenidos en las tablas de especificaciones de las mediciones, los que responden a la naturaleza de los marcos de referencia, a los propósitos y a las condiciones contextuales de las evaluaciones. Recomendamos un formato de tabla de especificaciones integrada, en la que el elemento nuclear sea el conjunto de indicadores de evaluación. Su proceso de elaboración supone la exhaustiva revisión y análisis de los marcos de referencia disponibles, a partir de los cuales se deben establecer las áreas de contenido que serán abordadas por la evaluación y que dan la estructura a la tabla como dominios y subdominios temáticos. Al interior de ellos se despliegan los indicadores de evaluación, que deben formularse con el propósito de operacionalizar el objeto de medida, en referencia a las particulares áreas de contenido, a través de tareas posibles de observar mediante el dispositivo o instrumento de evaluación. Estas tareas establecen tanto los contenidos como las habilidades con las que los evaluados deberán operar tales contenidos, y serán la orientación directa para la construcción de las preguntas y reactivos.

- Por el papel de puente que las tablas de especificaciones cumplen entre los objetivos de una evaluación y la información que finalmente será levantada -y por tanto, en el

resguardo de la validez en el uso de sus resultados-, consideramos que su desarrollo es un desafío técnico que debe enfrentarse sin escatimar recursos profesionales ni tiempo. En el caso de evaluaciones complejas y de altas consecuencias es necesario también contar con la participación y anuencia de las autoridades responsables; establecer qué y cuánto será evaluado tiene una dimensión política que no se limita a decisiones técnicas y que es imprescindible para la comunicabilidad, transparencia y legitimación de los procesos evaluativos.

Referencias

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, & NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION. (2014). *Standards for Educational and Psychological Testing.* Washington: American Educational Research Association.

ARGÜELLES, M., Castillo, M. Á. y Saragoni, C. (2015). Sistema de evaluación e instrumentos de medición del programa AEP. En B. Rodríguez, J. Manzi, C. Peirano, R. González y D. Bravo (eds., pp. 24-39), *Reconociendo el mérito docente.* Santiago, Centro de Medición MIDE UC.

DANIELSON, C. (2011). *Competencias docentes: desarrollo, apoyo y evaluación*. Recuperado de http://www.facultadeducacion.ucr.ac.cr/recursos/docs/PREAL%2051.Danielson.pdf

DOWNING, S. (2006). Twelve steps for effective. En S. M. Downing y T. M. Haladyna (eds., pp.3-26), *Handbook of Test Development*. Mahwah: Lawrence Erlbaum Associates.

DOWNING, S.M. y Haladyna, T.M. (eds.) (2006). *Handbook of Test Development*. Mahwah, Lawrence Erlbaum Associates.

FÖRSTER, C. y Rojas-Barahona, C. A. (2008). Evaluación al interior del aula: Una mirada desde la validez, confiabilidad y objetividad. *Revista Pensamiento Educativo, 43*, 285-305.

INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN. (2005). *Manual técnico: Especificaciones de reactivos.* Ciudad de México: Instituto Nacional para la Evaluación de la Educación. Recuperado de https://www.inee.edu.mx/index.php/bases-de-datos/bases-de-datos-excale/marcos-de-referencia/95-publicaciones/de-pruebas-y-medicion-capitulos/471-manual-tecnico-especificaciones-de-reactivos

LINN, R. (2006). The standards for educational and psychological testing: Guidance in test development. En S. M. Downing y T. M. Haladyna (eds., pp.27-38), *Handbook of Test Development.* Mahwah, Lawrence Erlbaum Associates.

LOUREIRO, G. (2009). *Evaluación en el aula, currículo y evaluaciones externas.* Montevideo: PREAL. Recuperado de http://repositorio.minedu.gob.pe/bitstream/handle/123456789/5000/Evaluaci%C3%B3n%20en%20el%20aula%2c%20curr%C3%ADculo%20y%20evaluaciones%20externas.pdf?sequence=1&isAllowed=y

MARTÍNEZ, F. (2016). *La evaluación de docentes en educación básica. Una revisión de la experiencia internacional.* Ciudad de México: Instituto Nacional para la Evaluación de la Educación.

MESSICK, S. (1989). Validity. En R. L. Linn (ed.). *Educational Measurement* (3a ed., pp. 13-104). Nueva York: American Council on Education /Macmillan.

MINISTERIO DE EDUCACIÓN DE CHILE. (2008). *Marco para la buena enseñanza.* Recuperado de https://www.docentemas.cl/docs/MBE2008.pdf

MURILLO, J. (2008). Hacia un modelo de eficacia escolar. Estudio multinivel sobre los factores de eficacia en las escuelas españolas. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 6*(1), 4-28.

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA. (2016a). *Reporte técnico: Tercer estudio regional comparativo y explicativo, TERCE.* Recuperado de http://unesdoc.unesco.org/images/0024/002471/247123s.pdf

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA. (2016b). *Informe de resultados tercer estudio regional comparativo y explicativo: Logros de aprendizaje 2015.* Recuperado de http://unesdoc.unesco.org/images/0024/002435/243532S.pdf

ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS. (2016). *PISA 2015: Resultados clave. 2015.* Recuperado de https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf

ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS. (2017). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy.* París: OECD Publishing.

OSORIO-ALVAREZ, M. C. y Parra, L. (2015). La satisfacción escolar en los estudiantes del primer año de la carrera de Médico Cirujano. *Investigación en Educación Médica, 5*(17), 3-10. https://doi.org/10.1016/j.riem.2015.08.002

PÉREZ, G., Ruiz, G., Langford, P., & Prado, P. (2017). *Marco de referencia para la documentación de experiencias y metodologías relevantes de evaluación formativa de escuelas.* Ciudad de México: Instituto Nacional para la Evaluación de la Educación.

RAVELA, P. (2006). *Fichas didácticas: Para comprender las evaluaciones educativas.* Montevideo: PREAL.

RAVELA, P., Picaroni, B. y Loureiro, G. (2017). *¿Cómo mejorar la evaluación en el aula? Reflexiones y propuestas de trabajo para docentes.* Montevideo: Grupo Almagro Editores.

WEBB, N. (2006). Identifying content for student achievement tests. En S. M. Downing, y T. M. Haladyna (eds., pp. 155-180), *Handbook of test development.* Mahwah: Lawrence Erlbaum Associates.

Instituto Nacional para la Evaluación de la Educación
México

Centro UC
Medición - MIDE