

Evaluación del Aprendizaje: Una guía práctica para profesores

Dra. Yolanda Edith Leyva Barajas

Marzo, 2010

Uno de los obstáculos que con mayor frecuencia tienen las instituciones educativas para realizar un trabajo sistemático de evaluación del aprendizaje, consiste en la falta de una cultura de la evaluación entre los actores del proceso educativo -alumnos, profesores y directivos-, lo que conduce a prácticas evaluativas que nos son congruentes con los contenidos, enfoques y propósitos del modelo educativo, o bien son insuficientes o superficiales. Tales prácticas suceden en gran medida por la carencia de bases teóricas y técnicas en algunas de las fases del proceso evaluativo: la planeación, el diseño y desarrollo de instrumentos de evaluación; el análisis de la información, o la interpretación y uso de los resultados. Estas carencias se reflejan en la emisión de juicios de valor subjetivos, superficiales o sesgados que impactan de manera negativa en los estudiantes. Generalmente estas prácticas no resultan útiles para la toma de decisiones que tienen que llevar a cabo los diferentes actores en los distintos momentos del proceso educativo.

En este documento se presenta información pertinente para que los grupos académicos de las instituciones educativas encuentren una guía práctica que coadyuve al desarrollo de las competencias necesarias para llevar a cabo una evaluación del aprendizaje congruente con los enfoques actuales de la educación basada en competencias, y útil para reorientar su intervención. Inicia con un breve marco teórico acerca del concepto de evaluación, así como las dimensiones con base en las cuales se ha desarrollado una clasificación o tipología de la evaluación, aspectos que serán de utilidad como sustento a una práctica educativa más congruente y eficaz. También se abordan aspectos metodológicos necesarios para la planeación colegiada de la evaluación del aprendizaje, el diseño y elaboración de los instrumentos y técnicas para la recogida de información y algunas reflexiones acerca del análisis e interpretación de la información para garantizar que la emisión de juicios de valor sea válida y por lo tanto útil para la mejora del proceso educativo.

Esta información constituye también un apoyo para orientar la práctica educativa del personal docente y a nivel institucional contribuye en la mejora de los sistemas de planeación y actualización de los planes de estudio, ya que la evaluación objetiva del aprendizaje es una fuente de información importante para evaluar la eficacia e impacto no solo de la intervención didáctica del docente, sino de la eficacia y funcionalidad del plan y programas de estudio.

Yolanda E. Leyva Barajas

1. Introducción a la Evaluación del Aprendizaje

Para iniciar haremos una breve reflexión acerca del concepto de evaluación educativa. Muchos son los autores que han ido definiendo el concepto a lo largo de la historia, la evolución del concepto ha permitido su construcción, integrando los nuevos enfoques o definiciones que a su vez se han ido construyendo a partir de las diferencias, carencias y perspectivas paradigmáticas en ocasiones contrapuestas, conduciendo a la elaboración de concepciones evaluativas diferentes a lo largo de la historia. No es el objetivo del presente hacer una revisión exhaustiva de este proceso, para ello sugiero revisar la reseña realizada por Ramos, Perales y Pérez (2009).

En este trabajo nos limitaremos a señalar que la concepción ha sufrido una evolución hasta llegar a la que actualmente es reconocida por la comunidad de investigadores y académicos dedicados a la evaluación educativa a nivel internacional, y que tal evolución lógicamente a enriquecido de manera progresiva la teoría, la metodología y desde luego los principios éticos necesarios para garantizar el desarrollo de procesos evaluativos útiles y creíbles. También revisaremos algunas propuestas para la clasificación o tipología que nos permitan ubicar a la evaluación del aprendizaje como proceso sistemático ligado a la intervención educativa.

Una síntesis actualizada acerca de las diferentes tendencias y definiciones de lo que actualmente representa la evaluación educativa es la referida por el grupo de evaluación y medición (*GEM*) de la Universidad de Valencia: Se entiende por Evaluación... "un proceso sistemático de indagación y comprensión de la realidad educativa que pretende la emisión de un juicio de valor sobre la misma, orientado a la toma de decisiones y la mejora" (Jornet, 2009).

Ofrece además un análisis de esta definición, destacando los siguientes aspectos:

- Se trata de **un proceso sistemático**; es decir, la evaluación debe ser un proceso racionalmente planificado como parte del desarrollo de la enseñanza, de forma que no debe entenderse como algo aislado, ni improvisado, ni desconectado del diseño y desarrollo de la docencia.
- **...de indagación y comprensión de la realidad educativa**; en este sentido, el elemento fundamental radica en el acercamiento a la realidad para conocerla adecuadamente y comprenderla, de forma que no puede darse una evaluación de calidad si no se sustenta sobre un grado de comprensión suficiente de la situación educativa de lo evaluado.
- **...que pretende la emisión de un juicio de valor** sobre la misma; finalmente se requiere emitir un juicio de valor, basado en criterios objetivos u objetivables, se entiende que un elemento de objetivación imprescindible en el contexto de la

evaluación educativa es el consenso intersubjetivo que pueden manifestar expertos en educación acerca de la calidad de los fenómenos educativos evaluados.

▪ **...orientado a la toma de decisiones;** es la base necesaria para poder tomar decisiones –de cualquier tipo, sean de mejora (evaluación formativa) o de rendición de cuentas (evaluación sumativa)-.

▪ **...y la mejora;** sólo puede entenderse que una evaluación es de calidad, si permite identificar no sólo los elementos que requieren mejora, sino el cómo dinamizar el proceso de mejora o innovación, es decir, el carácter formativo se identifica como un componente fundamental para cualquier evaluación.

En cuanto a las dimensiones con base en las cuales se clasifica a la evaluación educativa, retomamos la referida por Jornet, Sánchez y Leyva (2009) la cual se sitúa en el análisis de las cuestiones clave que guían cualquier plan de evaluación, y que por ello, orientan los procesos de la planificación y diseño. Esta clasificación, aunque originalmente fue realizada con objeto de identificar una tipología de planes de evaluación de docencia de educación superior, ha sido aplicada y adaptada a diferentes situaciones, es generalizable y se adapta al esquema conceptual anteriormente expuesto.

La propuesta incluye otras dimensiones que hemos denominado transversales las cuales se consideran complementarias porque refieren a las posibles especificaciones de los escenarios y situaciones educativas en los que se desarrolla la intervención educativa. De esta clasificación retomaremos por ahora, la dimensión que hace referencia a la unidad de análisis a la cual se dirige el proceso evaluativo. ¿Qué se evalúa o pretende evaluar? Las unidades identificables como referencia de los diversos procesos son:

- ❑ **Las personas.** En este epígrafe se pueden diferenciar evaluaciones referidas a alumnos o estudiantes –de cualquier nivel educativo-, profesorado y figuras o agentes de la educación (directores escolares, coordinadores académicos, orientadores educativos, supervisores...).
- ❑ **Las organizaciones educativas.** Las cuales se pueden clasificar como formales (sistemas educativos, niveles y/o modalidades educativas, instituciones escolares –de educación básica, media superior o superior); no formales (programas sociopolíticos de actuación formativa laboral, social, económica, empresarial), instituciones de formación para la empresa, entre otras.
- ❑ **Los elementos materiales de la intervención educativa.** Se incluyen aquí todos los elementos en que se materializa la organización educativa, tales como los proyectos educativos, los currícula o planes de estudio, programas de intervención educativa a cualquier nivel tanto del ámbito educativo como empresarial o social; y materiales educativos (libros de texto, fichas didácticas, material multimedia...).

Para el propósito de este trabajo, nos ubicamos en la evaluación de personas, de manera específica en la evaluación del aprendizaje de alumnos o estudiantes, esta primera definición nos brinda un primer paso para establecer la planeación del proceso evaluativo ya que orienta en gran medida la metodología de trabajo. Ahora bien se requiere contestar otras preguntas tales como ¿qué evaluar?, es decir qué aspectos o rasgos del aprendizaje vamos evaluar; el ¿para qué?, es decir, la finalidad y con ello el uso que se hará de los resultados de la evaluación. La definición del objetivo de la evaluación guiará otras fases de la planeación y desarrollo del proceso, tales como el diseño y desarrollo de técnicas e instrumentos de obtención y análisis de la información, así como el establecimiento de un marco que permita interpretar esta información, para dar sustento a la emisión de juicios y por lo tanto a la toma de decisiones más adecuadas.

La elección de un tipo determinado de evaluación depende de los objetivos de la misma, los cuales a su vez son subsidiarios de una perspectiva psicoeducativa, en virtud de la cual la evaluación adquiere un sentido específico (Miras y Solé, 1991). El marco psicoeducativo que se toma como referencia para interpretar la enseñanza y el aprendizaje, determina en gran medida tanto la información que es necesario obtener, como el tipo de juicios que se emitan y la naturaleza de las decisiones que se adopten.

El enfoque actual que hemos revisado del concepto de evaluación, en el caso específico de la evaluación del aprendizaje, implica que no debemos limitarnos al ámbito de los objetivos más fácilmente contrastables, porque se perdería de vista la importante influencia de estrategias intelectuales y de procesos de pensamiento subyacentes a las tareas específicas, factores que nos permiten explicar causalmente el origen del éxito o fracaso del alumno orientando la intervención apropiada a cada causa detectada; sólo así la evaluación cumple su función de perfeccionamiento y mejora del proceso educativo.

Para determinar la finalidad de un proceso de evaluación de aprendizaje, revisaremos ahora algunos conceptos que han dado origen a una tipología de la evaluación del aprendizaje, los cuales nos servirán para orientar la selección de técnicas y la construcción de instrumentos que recuperen la información necesaria y suficiente para sustentar el proceso evaluativo.

En el esquema presentado en la figura 1, Casanova (1998, 2007) representa la clasificación de la evaluación del aprendizaje en función de dimensiones tales como la funcionalidad u objetivo de la misma, la temporalidad en la cual se sitúa el proceso, el normotipo o tipos de referencias para la interpretación de los resultados de la evaluación; y finalmente la dimensión de los agentes, es decir quiénes van a emitir los juicios de valor.

El normotipo se refiere a la clasificación de los tipos de referencias o criterios que subyacen a las interpretaciones, en términos de si tales referencias o criterios son externos o internos al propio estudiante o persona sujeta a evaluación. La evaluación

nomotética, cuyo referente es externo, puede ser referida a una norma la cual depende de la población o grupo del cual forma parte el sujeto evaluado, o bien a un criterio establecido en algún programa educativo formal. Cuando la referencia es **normativa**, nos permite la comparación del rendimiento de cada alumno con respecto al logrado por el grupo, entonces es común escuchar que tal alumno tiene un rendimiento superior al rendimiento promedio de su grupo; o bien que un grupo específico está una desviación estándar por debajo del promedio nacional o estatal, ya que generalmente la norma representa la media de la población evaluada y los puntajes de cada persona o grupo evaluado pueden distribuirse en torno a esta media en una distribución normal. Si la referencia empleada es **criteria**, la interpretación es independiente del nivel logrado por el grupo o población a la cual pertenece el alumno o persona evaluada, en este caso el interés se centra en determinar el nivel de competencia del alumno con respecto a uno o más estándares previamente fijados.

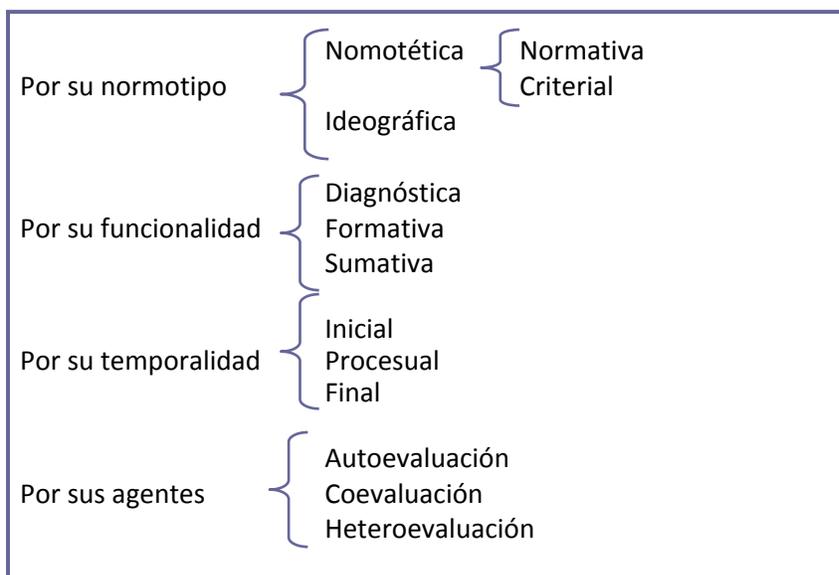


Figura 1. Tipología de la evaluación del aprendizaje

Por otra parte si lo que se requiere es conocer la evolución de cada alumno a lo largo de su proceso educativo, el criterio contra el cual comparar su ejecución es interno y es lo que se denomina evaluación **ideográfica**; la cual nos permite determinar los progresos que ha realizado cada alumno con respecto a su propia ejecución, en otros cursos o momentos de un curso y comparar con ellos las nuevas adquisiciones. Esta historia personal o individualizada resulta esclarecedora respecto a las causas de cada situación, en un momento dado y permite proyectar su posible evolución en el futuro, así como diseñar un proceso de intervención adaptado y con garantías de efectividad.

Los ejemplos de su utilidad son especialmente evidentes en programas de educación especial, o de preescolar; también en programas de educación física en escuelas en donde cada alumno presenta diferencias muy significativas en cuanto a

niveles mínimos de condición física presentados para poder iniciar el programa. En cada caso será necesario establecer un proceso fino de evaluación diagnóstica para determinar los niveles actuales de rendimiento y con base en ellos establecer un plan de intervención acorde a las capacidades de cada alumno para después en un proceso de evaluación continua comparar el rendimiento del alumno longitudinalmente verificando sus avances sin comparar su desempeño contra criterios externos a él mismo.

La dimensión de **funcionalidad** se refiere a la función que cumplirá el proceso evaluativo respecto al alumno, determina el uso que se hará de los resultados del mismo, generalmente en la bibliografía se encuentra ligada a la dimensión de **temporalidad**. De acuerdo con la naturaleza de la evaluación como actividad proyectada sobre todos los componentes del proceso didáctico, las decisiones tanto iniciales o diagnósticas, como continuo-formativas y finales, se proyectarán sobre la misma función docente, la metodología, los recursos y las circunstancias contextuales. Sólo a través de la evaluación sistemáticamente utilizada y aprovechada es factible transitar con cierta seguridad tanto en los procesos de planeación como de desarrollo del proceso educativo. La evaluación se relaciona con la planeación y la ejecución como parte de un todo absolutamente integrado.

La evaluación **inicial o diagnóstica** proporciona información acerca de los conocimientos y las habilidades previas del sujeto. Debe considerarse siempre en términos de su utilidad para facilitar adaptaciones constructivas de los programas educativos a los individuos. En este sentido cumple con la función reguladora que asegure que las características del sistema se ajusten a las de las personas a quienes se dirige; es decir que va a servir de base para adoptar decisiones relativas a la realización de actividades de apoyo, específicamente orientadas a la superación de problemas que presente el alumno, o bien dará lugar a modificaciones específicas en otros componentes de la enseñanza en función de los problemas detectados.

La evaluación formativa o procesual cumple una función reguladora de los procesos de enseñanza y de aprendizaje lo cual nos permite llevar a cabo ajustes y adaptaciones de manera progresiva durante el curso porque se centra más que en los resultados del aprendizaje en los procesos que se ponen en juego para el logro de tales resultados. Sólo centrados en los procesos podremos identificar áreas de oportunidad para poder ofrecer una retroalimentación apropiada a los estudiantes, de manera que ellos sepan qué es aquello que deben hacer o ajustar de su proceso para alcanzar los resultados esperados. Una estrategia de evaluación formativa requiere un marco conceptual que precise y justifique los:

- Los procesos implicados en la solución de diversas tareas (análisis de tareas)
- Las dificultades presentadas por la mayoría de los alumnos en el aprendizaje
- Métodos y técnicas para obtener información de los procesos implicados (métodos cualitativos de recogida de información)

- Principios de interpretación de esta información mediante marcos teóricos apropiados
- Estrategias de retroalimentación oportuna a los alumnos
- Pasos a seguir en la adaptación de estos procesos de aprendizaje

La evaluación **final o sumativa** constituye un balance general de los conocimientos adquiridos o de las competencias desarrolladas después de que se ha operado un programa de intervención educativa, se centra en los resultados del aprendizaje; es decir que se orienta a verificar el cumplimiento de los objetivos y estándares previamente determinados en el programa, y por lo tanto nos permite emitir un juicio de acreditación académica. Tal función de control, obliga a que el tipo de evaluación sea criterial en cuanto a su normotipo. Adicionalmente posibilita la comparación de grupos y la conexión entre niveles secuenciados, por lo que exige un espectro amplio de conductas representativas y significativas del aprendizaje alcanzado. La información cuantitativa proporcionada por la evaluación sumativa, permite además cumplir con la función de control y acreditación del aprendizaje, imprescindible para la planeación y sistematización de la práctica educativa.

Una reflexión relativa a la dimensión de funcionalidad es la relativa a los procesos mixtos; es decir, en los que se combinan consecuencias de ambos tipos en la evaluación - se identifican recomendaciones de mejora, al tiempo que se utiliza la información para apoyar una decisión de carácter sumativa-. Cualquiera de las finalidades implica toma de decisiones en cuanto a la intervención educativa. Desde la diferenciación realizada por Scriven (1967), los extremos de esta dimensión sumativo vs. formativo, se han tomado como elementos excluyentes. Desde la posición de Jornet, Sánchez y Leyva (2008), no se pueden entender totalmente así. Si asumimos a la educación como un proceso que se da a lo largo de toda la vida de la persona y en diferentes escenarios y situaciones educativas, es cierto que la intervención educativa estará acompañada por procesos evaluativos en los que se producen tanto actos orientados a la rendición de cuentas como otros dirigidos a la mejora.

En relación con la tipología en función de **los agentes**, Casanova (2007) refiere que la **autoevaluación** se produce cuando el sujeto evalúa sus propias actuaciones, es un tipo de evaluación que toda persona realiza a lo largo de su vida; en el caso que nos ocupa, es de suma importancia que el alumno realice de manera continua ejercicios de valoración de su aprendizaje, de manera que le sea posible identificar aspectos que debe mejorar. En la medida en que un alumno logre contrastar sus avances contra estándares de actuación establecidos, podrá identificar áreas de mejora, con lo cual estará en condiciones de regular su aprendizaje hacia el logro de competencias útiles para su desarrollo social y profesional.

La **coevaluación**, la describe como la evaluación mutua, conjunta de una actividad o trabajo determinado realizado entre varios. En este caso, lo recomendable es que

después de una serie de actividades didácticas, los participantes tanto alumnos como el profesor evalúen ciertos aspectos que consideren importantes de tal actuación conjunta. Generalmente tras un trabajo en equipos, de manera natural, cada uno valora lo que le ha parecido más interesante de los otros, por ejemplo se puede valorar si las actividades resultaron atractivas, si el contenido del trabajo realizado es pertinente, si el nivel de colaboración facilitó el logro de los objetivos, etc.; es muy importante en la conducción de estos procesos de coevaluación pedir a los alumnos que se centren en la valoración tanto de los aspectos positivos o que ellos consideren como los más destacados, como en aquellos que es necesario trabajar más para mejorar la calidad del trabajo desarrollado en conjunto.

La **heteroevaluación** consiste en la evaluación que realiza una persona sobre el trabajo, actuación o rendimiento de otra persona. Es aquella que habitualmente hace el profesor de sus alumnos. Dado que es un proceso importante e imprescindible de control en los esquemas y modelos educativos vigentes, rico por los datos y posibilidades que ofrece, delicado por el impacto que tiene en las personas evaluadas, y complejo por las dificultades técnicas que supone la emisión de juicios de valor válidos y objetivos; es que estamos proponiendo esta guía práctica para profesores.

Como nota complementaria a esta propuesta de tipología de Casanova (2007) es importante señalar que los elementos sistémicos y holistas que se integran en la concepción de la educación, apuntan a que **la finalidad** de los procesos puede ser diferente según se analice desde una u otra unidad de análisis. Por poner un ejemplo; una prueba de selección o admisión constituye sin duda alguna una evaluación **sumativa** para la persona que debe superarla, pero –desde determinados modelos de gestión educativa– puede entenderse como un acto orientado a la selección de estudiantes con el fin de asegurar unos niveles mínimos de competencia en aquellos que son admitidos, de manera que puede formar parte de una orientación dirigida a la mejora de la gestión de la institución. Este ejemplo ilustra que la finalidad depende del plano de la intervención desde el que se analice, en relación a la unidad a que se impute (estudiante o institución), y del modelo de gestión que la utilice. Por ello, la catalogación en los extremos de este continuo siempre se debe entender referida a la unidad objeto de análisis y como un indicador de la tendencia dominante en la finalidad del proceso evaluativo.

Una recomendación especialmente útil para la planeación del proceso evaluativo, es el reconocimiento de lo que Jornet y Leyva (2008) definen como las dimensiones transversales, ya que son las que identifican características del escenario y situación educativa en que se desarrolla la intervención, tales dimensiones se refieren al grado o nivel educativo, la edad de los estudiantes (adultos, adolescentes o niños); la modalidad de entrega -presencial o en línea-; entre otros. La multidisciplinariedad, junto a la variabilidad y diversidad del objeto de estudio en ciencias de la educación, y la evidencia puesta de manifiesto por las necesidades sociales acerca de la utilidad de la intervención,

han ido abriendo los marcos de referencia en que sustentar el análisis de las diversas unidades objeto de evaluación.

La comprensión de la realidad con el fin de intervenir sobre ella, requiere de aproximaciones muy diferenciadas en cuanto al grado de formalización del conocimiento. En el caso de la evaluación del aprendizaje, a un nivel macroanalítico, existe un mayor desarrollo teórico que en el de las organizaciones educativas. Asimismo, hay zonas disciplinares de la evaluación de la educación, en cada una de las unidades objeto, que tienen mayor tradición y desarrollo que otras. Por ejemplo, el desarrollo teórico en el ámbito de las aptitudes es mayor que en el de las habilidades sociales o que en el rendimiento. Este desarrollo teórico-metodológico desigual -en un momento en el que ha sido muy importante la penetración de los movimientos de calidad en diversas áreas de las Ciencias Sociales-, así como un planteamiento epistemológico más flexible, ha tenido como consecuencia que los referentes acerca de la calidad de los procesos de intervención se incluyan como elementos que orientan la identificación de los criterios de interpretación de los procesos evaluativos (Jornet, Sánchez y Leyva, 2008).

A manera de resumen de este muy breve marco teórico, evaluar el proceso educativo que implica a la enseñanza y al aprendizaje, significa interpretar los resultados obtenidos de la actividad que conjunta a profesores y alumnos en cuanto al logro de los objetivos de la educación, sean éstos referidos a conocimientos o al desarrollo de competencias, además de que la evaluación del aprendizaje como proceso sistemático mediante el cual se recoge información acerca del aprendizaje del alumno, permite no sólo mejorar ese aprendizaje y detectar elementos para formular un juicio acerca del nivel alcanzado, sino que también informa acerca de la calidad del aprendizaje logrado y de lo que el alumno es capaz de hacer con ese aprendizaje, aspecto especialmente útil para el enfoque orientado al desarrollo de competencias.

Como proceso sistemático podemos considerar a la evaluación como una serie de actividades planeadas con suficiente anticipación que responden a intenciones claras y explícitas y que guardan una relación estrecha y específica con el programa escolar, con las actividades de enseñanza y de aprendizaje y con las circunstancias en que se dan esas actividades. Las características del aprendizaje, las posibilidades del profesor y las circunstancias en que ha de desarrollarse la evaluación, determinan cual es la estrategia más adecuada, o simplemente posible, y a que nivel de rigor o complejidad puede llegarse. Lo que es realmente indispensable, es que el juicio de evaluación no se base en impresiones subjetivas, situaciones excepcionales o aisladas, ni en información insuficiente, irrelevante o accesorio.

Además de la estrategia de recogida y análisis de la información, los cuerpos colegiados o academias deben establecer los estándares de ejecución esperados como resultado del análisis y discusión de lo que persiguen los programas de intervención, facilitando con ello una interpretación adecuada del nivel logrado por los estudiantes. Los

estándares deben ser difundidos entre los estudiantes, de manera que el proceso de evaluación sea transparente, justo y equitativo, sólo así podrá ser útil para apoyar la toma de decisiones de todos los usuarios: los propios estudiantes, los profesores y académicos tanto a nivel individual como colegiado, y para aquellos que asumen actividades de dirección en los centros escolares, lo cual en suma representa “la creación de una cultura de la evaluación”.

2. Planeación de la Evaluación del Aprendizaje

Resulta evidente bajo la perspectiva anteriormente señalada, que un proceso evaluativo requiere de una fase de planeación la cual de acuerdo con estándares de calidad en la elaboración de pruebas educativas y psicológicas (APA, 1999) debe ser realizada por academias o grupos colegiados de profesores de una institución educativa ya sea que estos grupos se organicen por materia, semestre o por área curricular específica del plan de estudios. Un proceso de planeación de la evaluación requiere que tales academias o grupos colegiados conozcan el perfil de egreso y el plan de estudios o estructura curricular que guía y organiza las diversas materias o asignaturas, ya que deberán realizar un análisis de cómo contribuyen los programas de tales materias o asignaturas al logro de las competencias expresadas en el perfil de egreso.

Este ejercicio de discusión y análisis colegiado servirá para llegar a acuerdos respecto a qué evaluar, es decir que deberán decidir cual es la información de desempeño de los estudiantes que sirva de sustento para hacer inferencias acerca de los procesos y resultados de aprendizaje esperados, así como los niveles mínimos de desempeño o competencia que se exigirán de los alumnos. Para una revisión de lo que implica la evaluación de competencias revisar el apunte de Evaluación de Competencias de Leyva (2009). Una vez definido el qué evaluar, deberán decidir el para qué evaluar, es decir la finalidad de la evaluación y el uso que se hará de los resultados del proceso, también deberán llegar a acuerdos acerca de cuáles son las técnicas e instrumentos más apropiados de recogida y análisis de la información y de cómo conviene interpretar los resultados.

Antes de iniciar cualquier acción, es indispensable hacer las especificaciones pertinentes, para después seleccionar los materiales que tengan congruencia con el plan preconcebido. El desarrollo del esquema requiere una detallada consideración de los objetivos, enfoques y contenidos de las unidades de instrucción que contienen los programas de estudio. Un procedimiento recomendado por Gronlund (1982) es el de desarrollar una tabla de especificaciones, seguida de amplias categorías que pueden ser subdivididas de acuerdo con la naturaleza de la instrucción y la complejidad de sus finalidades, para asegurarnos que el campo ha sido adecuadamente considerado tanto en los contenidos como en el tipo de conducta que se deberá manifestar o utilizar, dado que de esta fase del proceso depende la validez de constructo y contenido del proceso evaluativo.

La tabla de especificaciones reúne información que nos permite verificar la congruencia interna del proceso evaluativo; ya que como resultado del análisis y discusión colegiada de los profesores, se redactarán indicadores de aprendizaje que en su conjunto reúnan la evidencia suficiente del desarrollo de competencias y del grado de cumplimiento de los objetivos o propósitos expresados en los programas de estudio de las

asignaturas. Una expresión adecuada del enunciado de cada indicador de aprendizaje, debe comprender tres elementos: a) la conducta o ejecución específica que debe exhibir el estudiante, la cual se relaciona directamente con el proceso o resultado de aprendizaje a inferir; b) las condiciones en que deberá manifestarse tal conducta o ejecución, y c) el estándar o norma de ejecución aceptable.

La educación formal, se guía por procesos claramente organizados por un currículo o plan de estudios y programas de estudio de las asignaturas o materias que integran las diferentes líneas curriculares del plan; por lo que podemos partir de los objetivos y de los contenidos de cada una de las unidades de instrucción desglosadas por temas y subtemas en el caso de planes diseñados por objetivos; o bien las competencias a lograr junto con sus componentes declarativos, procedimentales y actitudinales. Cuando estos no existen o no están actualizados, se debe iniciar desde un perfil referencial de validez, con el subsecuente desglose de objetivos y contenidos necesarios para especificar lo que se desea evaluar.

Una forma apropiada de hacer un plan detallado de evaluación o tabla de especificaciones consiste en redactar, por cada unidad temática del programa, o competencia, una serie de indicadores específicos que se aceptarán como evidencia de que se han logrado los resultados de aprendizaje propuestos, en el nivel de profundidad deseado. Para elaborar esta tabla de especificaciones es conveniente el uso de algún sistema de clasificación de tales indicadores de aprendizaje de acuerdo con el grado de profundidad y alcance que se espera lograr en cada programa, cuidando que el enfoque y propósito general de la materia o asignatura en cuestión, quede adecuadamente representado, para lo cual se recomienda el uso de una taxonomía.

Para esta parte del diseño se recomienda una taxonomía de objetivos educativos acorde con los descubrimientos más recientes en materia de procesos implicados en el aprendizaje. Para tal propósito se realizó una revisión y análisis de diversas taxonomías en la literatura especializada. Entre las más conocidas se encuentran las publicadas por Bloom (1956), Gagné (1972), Tuckman (1979), Merrill (1994), Haladyna (1997), Marzano (2000, 2006); aunque existen otros enfoques diferentes o complementarios. Como resultado de tal revisión, para esta guía práctica propongo utilizar la taxonomía de Marzano (2000, 2006) la cual ha resultado ser útil y congruente con los modelos educativos actuales basados en competencias tanto de nivel medio superior como en diversas carreras de educación superior.

Marzano desarrolla su propuesta, a partir de un análisis de la taxonomía de Bloom, principalmente de sus debilidades. Para su integración, rescata aquellos aspectos aún vigentes de la taxonomía publicada por Bloom en 1956, incorporando los avances de los más de 40 años posteriores a su publicación. Cabe hacer un paréntesis y señalar cuáles son las críticas más comunes a la taxonomía de Bloom, y en que consisten; una de ellas es que se sustenta en una concepción del aprendizaje a un nivel unidimensional, asumiendo el constructo de *complejidad* como único elemento para separar los niveles del dominio

cognoscitivo, otra crítica relacionada con este constructo, es acerca de la estructura jerárquica que presupone que aquellos niveles subordinados involucran menor nivel de complejidad de los procesos cognoscitivos inherentes. Ambas características de la taxonomía no se pueden sostener desde los descubrimientos actuales de investigación básica de los procesos cognoscitivos en el hombre (Furst, 1994).

Estas críticas se sustentan en el principio básico de que la dificultad de un proceso mental es una función de al menos dos factores: la complejidad inherente al proceso en términos de los pasos involucrados en el procesamiento de la información y en el grado de familiaridad con el proceso. La complejidad de un proceso mental es invariante, es decir que el número de pasos y sus relaciones no cambian, no obstante la familiaridad con el proceso puede cambiar a lo largo del tiempo. A mayor familiaridad con un proceso se asocia un menor tiempo de ejecución y una mayor facilidad en la misma. En consecuencia no es conveniente ordenar jerárquicamente los procesos mentales en función de su complejidad, resulta más efectivo ordenarlos en términos de control, ya que actualmente sabemos que algunos procesos ejercen un control sobre la operación de otros procesos.

El modelo teórico en que se sustenta la taxonomía, es propiamente una teoría del pensamiento humano, y no una estructura como en el caso del sustento de la taxonomía de Bloom. Técnicamente los modelos y las teorías son sistemas que permiten predecir fenómenos, mientras que una estructura representa conjuntos organizados de principios que describen características de un fenómeno dado pero no necesariamente permiten la predicción del fenómeno. La propuesta de Marzano (2000, 2006) consta de tres sistemas de conocimiento interrelacionados: *self-system* o sistema de autoconciencia, que integra autoeficacia y motivación; el sistema metacognitivo y el sistema cognoscitivo.

- a) El sistema denominado *self-system* consiste en una red de creencias y metas interconectadas que permiten emitir juicios acerca de la conveniencia de involucrarse en una nueva tarea. Este sistema es también un primer determinante de la motivación que uno desarrolla frente a la tarea. Si una tarea se considera como relevante o importante, si la probabilidad de éxito para resolverla es alta y si existe algún efecto positivo asociado a ella, el individuo estará motivado a comprometerse con la tarea.
- b) El sistema metacognitivo actúa una vez que la tarea se ha seleccionado, es responsable de establecer metas relacionadas con la nueva información y de diseñar estrategias para alcanzar las metas. Una vez activado el sistema metacognitivo interactúa de manera continua con el sistema cognoscitivo.
- c) El sistema cognoscitivo cuya función es propiamente el procesamiento de información fundamental para el logro de la tarea, es el responsable de las operaciones analíticas tales como hacer inferencias, comparar y clasificar información entre otras.

Esta teoría de aprendizaje, permite el diseño de un sistema jerárquico del pensamiento humano desde la perspectiva de dos criterios: el flujo de información y el

nivel de control de los procesos involucrados. En términos del flujo de información, el procesamiento de la información siempre inicia con el sistema de autoeficacia, prosigue con el sistema metacognitivo y finaliza con el sistema cognoscitivo, es decir que los tres sistemas representan una organización jerárquica en términos del flujo de la información. Además de que el estatus de los diversos factores dentro de un sistema controla el estatus de los diferentes factores dentro de los demás sistemas.

Por ejemplo si dentro del sistema de autoeficacia no existen creencias que concedan importancia a la tarea, o se piensa que tiene pocas probabilidades de lograr una ejecución aceptable, el individuo no se involucrará en la tarea o su nivel de compromiso será mínimo o de baja motivación. Además, aún si la tarea se percibe como importante y el sujeto piensa que puede realizarla, pero el sistema metacognitivo no establece metas claras o no existe un monitoreo efectivo, la ejecución de la tarea puede fallar. Finalmente aún cuando funcionen bien todas las operaciones de los dos primeros sistemas, pero el procesamiento de la información dentro del sistema cognoscitivo no opera efectivamente, la tarea no se completará.

Adicionalmente estos tres sistemas también tienen una jerarquía relativa al nivel de conciencia requerida o recursos de procesamiento consciente para controlar la ejecución. Es decir que en el sistema de autoeficacia los procesos para establecer la relevancia, eficacia y motivación hacia la ejecución de una tarea, representan un nivel de introspección y pensamiento consciente mayor que aquel involucrado en los procesos metacognitivos de establecimiento de metas y monitoreo consciente. Finalmente en el sistema cognoscitivo, los procesos de utilización de la información son los que requieren mayor grado de pensamiento consciente en su ejecución, seguidos de los de análisis, comprensión y por último los de recuperación que pueden darse de manera automática.

En el caso de la elaboración de pruebas educativas realizadas a partir de reactivos de opción múltiple, para evaluación sumativa, una vez que se redacten los indicadores de aprendizaje acordes a los objetivos o competencias que se desea evaluar, se deben realizar ponderaciones por área, ya que la tabla de especificaciones cumple una doble función: a) sirve de guía a los redactores de reactivos acerca de la naturaleza de la evidencia que se desea recuperar para verificar si el sujeto evaluado tiene el nivel requerido de conocimientos o competencias y b) permite obtener validez de contenido y constructo al verificar la congruencia entre los reactivos elaborados y los indicadores de aprendizaje planteados. Adicionalmente la ponderación nos indica el esquema de muestreo más apropiado para evaluar el universo de conocimientos y habilidades con base en un análisis de relevancia, garantizando que tal muestreo sea representativo para poder emitir juicios válidos a partir de los resultados de la aplicación de las pruebas.

En la **tabla 1**, se presenta un resumen de la taxonomía que Marzano publicó en 2000 la cual mantiene en la segunda edición publicada en 2006, con una descripción de cada uno de los tres sistemas y por cada nivel de estos sistemas.

**Tabla 1. Taxonomía de Objetivos Educativos
Los tres sistemas de pensamiento**

	Descripción de los objetivos
Nivel 6: Self-System	
Examinando relevancia	Identifica la importancia del conocimiento y el razonamiento que subyace a su percepción
Examinando eficacia	Identifica las creencias acerca de su habilidad para mejorar su competencia o comprensión del conocimiento y el razonamiento que subyace a esta percepción
Examinando respuesta emocional	Identifica respuestas emocionales al conocimiento y las razones de estas respuestas
Examinando motivación	Identifica su nivel de motivación para mejorar su competencia o comprensión relativa al conocimiento y las razones de este nivel de motivación
Nivel 5: Sistema de Metacognición	
Especificación de metas	Establece un plan de metas relativas al conocimiento
Monitoreo de procesos	Monitorea la ejecución del conocimiento
Monitoreo de Claridad	Determina el grado en el cual tiene claridad acerca del conocimiento
Monitoreo de exactitud	Determina el grado de exactitud del conocimiento
Sistema Cognoscitivo	
Nivel 4: Utilización del conocimiento	
Toma de decisiones	Usa su conocimiento para tomar decisiones o puede tomar decisiones acerca del uso de su conocimiento
Solución de problemas	Usa su conocimiento para resolver problemas o puede resolver problemas acerca de su conocimiento
Cuestionamiento experimental	Usa su conocimiento para generar y probar hipótesis o genera y prueba hipótesis acerca de su conocimiento
Investigación	Usa su conocimiento para conducir investigaciones o puede conducir investigaciones acerca del conocimiento
Nivel 3: Análisis	
Empate	Identifica importantes similitudes y diferencias entre el conocimiento
Clasificación	Identifica categorías superordinadas y subordinadas relacionadas con el conocimiento
Errores de análisis	Identifica errores en la presentación o uso de la información
Generalización	Construye nuevas generalizaciones o principios basados en el conocimiento
Especificación	Identifica aplicaciones específicas o consecuencias lógicas del conocimiento
Nivel 2: Comprensión	
Síntesis	Identifica la estructura base del conocimiento y las características o componentes críticos de los que no lo son. Extrae la esencia de la información generando una macroestructura.
Representación	Reconoce características de información y crea un análogo simbólico de la información contenida en una macroestructura. La información puede ser procesada en dos modos primarios: lingüístico o de imágenes.
Nivel 1: Recuperación	
Recuerdo	Reconoce características de información, aunque no necesariamente entiende la estructura del conocimiento o puede diferenciar componentes críticos.
Ejecución	Ejecuta un procedimiento sin errores significativos pero no necesariamente entiende cómo y por qué funciona.

En esta propuesta taxonómica, a diferencia de la propuesta publicada por Bloom, Marzano divide de manera clara el dominio del conocimiento y las operaciones mentales, al postular tres sistemas de pensamiento que operan con tres dominios de conocimiento. En la **tabla 2**, se presenta una síntesis de tales dominios del conocimiento con sus componentes, una descripción para cada componente y un ejemplo. Es importante destacar que dentro de cada dominio, existe un orden jerárquico que tiene que ver con la complejidad de cada componente.

Tabla 2. Componentes de los tres dominios de conocimiento

Dominio de Información			
Componentes		Descripción	Ejemplos
Detalles	Vocabulario	Significado de términos a un nivel general y muchas veces superficial.	El significado del término “democracia” sin los detalles de sus implicaciones en diversos contextos.
	Hechos	Contienen información específica acerca de personas, lugares, cosas y eventos.	Las características de un lugar: “Chetumal es la capital de Quintana Roo”.
	Secuencias de tiempo	Eventos importantes que ocurren entre dos puntos en el tiempo.	Los eventos que ocurren entre el inicio de una campaña política y las elecciones presidenciales en un país.
	Secuencias causa-efecto	Involucran eventos que producen un producto o un efecto. Comúnmente los efectos tienen redes complejas de causas.	Organización de secuencias causales de algún evento histórico. Secuencia de los procesos de división celular.
	Episodios	Eventos específicos que ocurren en un determinado tiempo y lugar, entre personas específicas y con secuencias de tiempo y causales particulares.	Un debate político entre candidatos de los partidos políticos, previo a las elecciones presidenciales de un país.
Ideas organizadas	Generalizaciones	Son declaraciones de carácter general, representan abstracciones que identifican clases o categorías de personas, lugares o situaciones.	El amor es una de las emociones humanas más poderosas.
	Principios	Tipos específicos de generalizaciones que se refieren a relaciones causales o correlacionales.	Causal: La tuberculosis es producida por una bacteria denominada “ <i>tubercle bacillus</i> ” Correlacional: Las ciudades con mayor densidad de población presentan índices de criminalidad elevados.

Dominio de procedimientos mentales			
Componentes		Descripción	Ejemplos
Habilidades	Reglas simples	Es el tipo más simple, consiste de una producción “ <i>si</i> (una situación X)- <i>entonces</i> ejecuta la acción Y)” y se emplean en los pasos de algún procedimiento	El procedimiento para realizar una resta entre dos números naturales.
	Algoritmos	Procedimientos mentales que normalmente no presentan variantes en su aplicación, tienen resultados específicos y un número específico de pasos.	Obtener la media aritmética o la varianza de un grupo de datos.
	Tácticas	Consisten en la producción de un patrón general en el cual se incorporan reglas generales, no hay un orden o número de pasos rígido.	Interpretación de un gráfico que presenta una función entre variables.
Procesos	Procedimientos macro	Técnicamente se refieren a procedimientos más complejos con varios subcomponentes y que requieren alguna forma de administración y control.	Elaborar un plan de mercadotecnia de algún producto o servicio.
Dominio de procedimientos psicomotores			
Componentes		Descripción	Ejemplos
Habilidades	Procedimientos básicos o fundamentales	Estos procedimientos se desarrollan generalmente sin una instrucción formal, no obstante pueden mejorarse, por lo que pueden clasificarse como tipos de conocimiento en la medida en que son habilidades aprendidas.	Fuerza estática Equilibrio corporal Velocidad de movimientos Velocidad de digitación Destreza digital Destreza manual Estabilidad, uniformidad y consistencia brazo-mano Control, precisión
	Combinaciones simples de procedimientos	Involucra pasos de un procedimiento en el cual actúan procedimientos básicos en paralelo.	Lanzar un tiro libre en basquetbol
Procesos	Combinaciones complejas de procedimientos	Conjunto organizado de procedimientos complejos para el logro de una meta específica.	Extraer una pieza dental. Conducir un aeroplano en diversas condiciones climáticas y de visibilidad.

Así los procesos mentales que subyacen a los sistemas de pensamiento interactúan de diversas maneras con los tres dominios de conocimientos, los cuales pueden clasificarse en tres categorías generales: a) el dominio de la información -que se refiere al conocimiento declarativo-; y aquel que se conoce como conocimiento procedimental: b) el dominio de procedimientos mentales y c) el de los procedimientos psicomotores. Dado que estos tres tipos de conocimiento son muy diferentes, resulta conveniente pensar en ellos como dominios sobre los que actúan los tres sistemas del pensamiento.

Ejemplos:

Ejemplo 1. Tabla de especificaciones para evaluar habilidades lectoras

Competencia para la comunicación escrita.	Evidencia necesaria para verificar que se desarrolló la competencia a evaluar. Se deben enunciar tantos indicadores como el grupo colegiado determine que es suficiente.	Niveles de complejidad sugeridos						1. Recuperación 2. Comprensión 3. Análisis 4. Utilización del conocimiento 5. Metacognición 6. Autoeficacia	
		1	2	3	4	5	6	Dominio	Ponderación
Bloques o rasgos de competencias	Indicadores de aprendizaje								
Comprensión lectora de textos informativos, científicos y literarios tanto continuos como discontinuos.	Identifica las ideas centrales de diversos tipos de textos		X					S. Causa-efecto	50%
	Identifica el propósito de diversos tipos de textos			X				Episodio	
	Diferencia los propósitos de diversos tipos de texto.			X				Generaliz.	
	Relaciona ciertas ideas contenidas en el texto para obtener información que le permite comprobar hipótesis.				X			Proced. macro	
	Elabora inferencias a partir de la comparación de distintas fuentes.				X			Proced. macro	
	Reconoce sus aciertos y sus fallas en la comprensión de diversos tipos de texto.					X		Proced. macro	
	Reconoce sus intereses y gustos entre los diversos tipos de textos.						X	Proced. macro	
Precisión en el uso del vocabulario y en el uso de las estructuras lingüísticas en la comunicación escrita.	Selecciona las palabras adecuadas de acuerdo al contexto		X					Tácticas	30%
	Identifica el significado de palabras y expresiones a partir del contexto		X					S. Causa-efecto	
	Aplica las metáforas de manera adecuada y acorde al contexto en diversas situaciones.			X				Proced. macro	
	Aplica adecuadamente las reglas de concordancia entre adjetivo y sustantivo y entre verbo y sustantivo				X			Algoritm.	
	Aplica los tiempos verbales correctamente de acuerdo con el mensaje a comunicar				X			Algoritm.	
	Detecta sus errores gramaticales y semánticos en sus escritos.					X		Proced. macro	
Redacción de textos informativos y científicos.	Completa oraciones que emiten un mensaje claro y específico.		X					S. Causa-efecto	20%
	Diferencia oraciones primarias y secundarias de un párrafo a partir de la intención comunicativa.			X				S. Causa-efecto	
	Establece nexos entre los párrafos de un texto con una intención comunicativa clara y explícita.			X				S. de tiempo	
	Jerarquiza en un orden lógico los elementos a comunicar (plan de redacción)				X			Proced. macro	
	Reconoce y corrige fallas detectadas en sus escritos.					X		Proced. macro	100%

Ejemplo 2. Tabla de especificaciones para evaluar un curso de matemáticas

Competencias en la solución de problemas mediante el uso de las Matemáticas	Evidencia necesaria para verificar que se desarrolló la competencia a evaluar. Se deben enunciar tantos indicadores como el grupo colegiado determine que es suficiente.	Niveles de complejidad sugeridos						1. Recuperación 2. Comprensión 3. Análisis 4. Utilización del conocimiento 5. Metacognición 6. Autoeficacia	
		1	2	3	4	5	6	Dominio	Ponderación
Bloques o rasgos de competencias	Indicadores de aprendizaje								
Uso de las operaciones numéricas básicas.	Aplica los algoritmos de las operaciones básicas en la solución de ejercicios con números enteros y fraccionarios.	X						Algoritmo	20%
	Aplica la jerarquía de las operaciones básicas con enteros y fracciones.		X					Algoritmo	
	Encuentra el mínimo común múltiplo y el máximo común divisor.		X					Reglas simples	
	Identifica sus tipos de errores en la solución de operaciones con números enteros y fracciones.					X		Proced. macro	
Uso de las operaciones con potencias y raíces.	Aplica las leyes de los exponentes en operaciones aritméticas.		X					Algoritmo	20%
	Convierte de la notación de radical a la notación de potencia		X					Algoritmo	
	Aplica las leyes de los radicales		X					Reglas simples	
	Identifica y corrige sus errores en la aplicación de las leyes de los exponentes.					X		Algoritmo	
	Reconoce sus capacidades para el uso de operaciones con potencias y raíces.						X	Proced. macro	
Traducción y operaciones con polinomios	Traduce los polinomios del lenguaje común al algebraico y viceversa.		X					Tácticas	20%
	Establece en forma algebraica las relaciones entre datos a través de polinomios.			X				Tácticas	
	Identifica términos semejantes en los polinomios		X					Reglas simples	
	Identifica el tipo de errores que comete en la traducción y establecimiento de relaciones entre los polinomios.					X		Tácticas	
Solución de problemas mediante el uso de las matemáticas.	Plantea mediante el lenguaje algebraico un problema para su solución.			X				Tácticas	40%
	Establece relaciones entre variables de un problema mediante el lenguaje matemático.			X				Proced. macro	
	Selecciona alternativas viables de solución a problemas planteados.				X			Proced. macro	
	Soluciona problemas mediante el uso de las matemáticas.				X			Proced. macro	
	Corrige las soluciones dadas a diversos problemas planteados mediante la identificación del tipo de error que cometió.					X		Proced. macro	
	Identifica sus intereses y gustos por las matemáticas y la solución de problemas.						X	Proced. macro	

3. Técnicas e instrumentos de recogida de datos para la evaluación formativa

Para la *evaluación formativa*, se recomienda obtener toda la información necesaria para regular el proceso educativo en el aula, por lo que se sugiere seleccionar entre las diversas técnicas disponibles para recabar la información necesaria acerca del progreso y dificultades que los alumnos enfrentan *en el proceso de aprendizaje*, así como desarrollar los instrumentos que resulten más apropiados para registrar y analizar dicha información.

Casanova (1998, 2007) presenta un paradigma evaluador para la evaluación formativa centrado en técnicas cualitativas de recogida de información, el cual se representa en el esquema presentado en la figura 2.

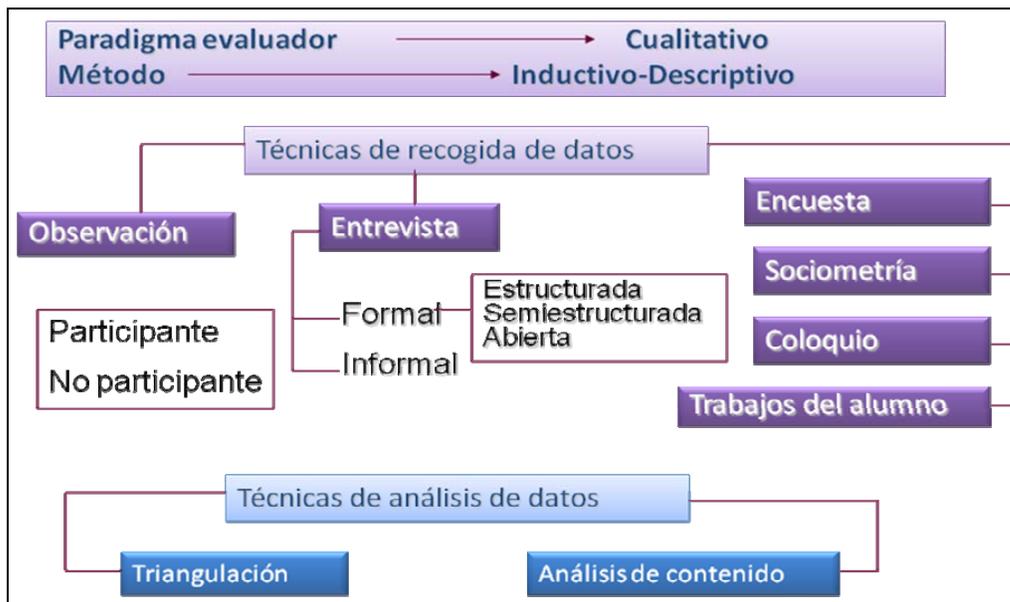


Figura 2. Paradigma de evaluación formativa

En el proceso de planeación de la evaluación una vez definido el propósito de la misma y los indicadores de aprendizaje a evaluar, se requiere decidir cuáles son las técnicas e instrumentos más apropiados a estos propósitos y al uso que se hará de los resultados del proceso evaluativo, si el propósito es formativo, se recomienda como lo señala Casanova (1998; 2007) un paradigma cualitativo que permita recoger información de los procesos, por lo que se requiere integrar diversas técnicas e instrumentos como parte de una estrategia que además permita la triangulación de fuentes de información para la interpretación y la inferencia de procesos de alto orden asociados al aprendizaje.

No se trata sólo de reunir mucha información, se requiere tener una estrategia que permita integrar la información en un todo coherente que apoye la toma de decisiones continua y oportuna para ir ajustando la intervención educativa hacia el logro de los propósitos o desarrollo de las competencias. Para plasmar la información obtenida mediante las técnicas que se apliquen, es muy importante contar con instrumentos o registros adecuados a cada técnica para dar garantía de rigurosidad y sistematicidad.

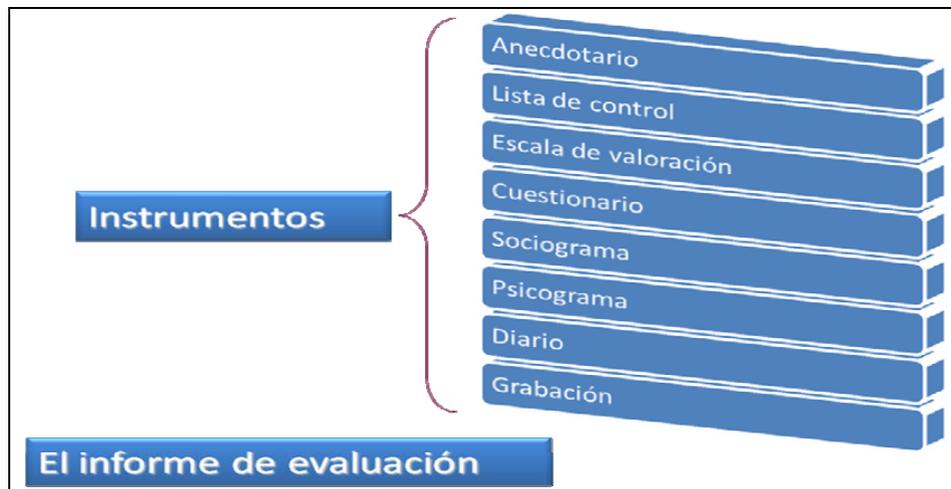


Figura 3. Instrumentos de registro de información

Conviene también una revisión a la propuesta de Tenbrink (1997) relativa a técnicas e instrumentos para la evaluación de aspectos del desempeño que requieren la observación de procedimientos psicomotores en tiempo real, o bien la evaluación de los productos derivados del desempeño de alguna actividad y para la evaluación de actitudes y otros rasgos socio-personales; aspectos implicados en el concepto de competencia (Leyva, 2009).

Los instrumentos más comunes que sirven para observar de manera sistemática el rendimiento y la ejecución de los estudiantes, así como aspectos o rasgos socio-personales y de actitudes, son las listas de control, también llamadas listas de cotejo y las escalas de evaluación.

Para la construcción de listas de control válidas, fiables y fáciles de usar se requiere:

1. Especificar una ejecución o producto adecuado
2. Enumerar las conductas o características más importantes para verificar si se producen
3. Añadir cualquier error común
4. Ordenar la lista de conductas o características en el orden en que se producen
5. Ofrecer un modo de utilizar la lista

Las escalas de evaluación, son variantes de las listas de control en el sentido de su finalidad y construcción; no obstante tienen la ventaja de que nos permiten formular juicios sistemáticos acerca del grado en el que se presenta algún comportamiento o característica relacionado con el desempeño a observar. Consisten en un conjunto de características a juzgar y algún tipo de jerarquía, y el observador usa la escala para indicar la cualidad, cantidad o nivel de rendimiento observado.

Ambos tipos de instrumentos son útiles porque ayudan al observador a centrarse en aspectos específicos, observables del elemento que se juzga, de manera que cuando se usa con un grupo de estudiantes, todos serán juzgados desde un marco de referencia común, lo que ofrece objetividad y confiabilidad; siempre y cuando se ponga cuidado en la enumeración de las características o elementos; y en el caso de las escalas de evaluación, la escala esté bien diseñada. Tenbrink (1997) ofrece algunas recomendaciones útiles para su construcción en términos de aquello que se pretende evaluar:

En el caso de la evaluación de procedimientos como cantar, tocar un instrumento, dibujar, dirigir una discusión, debatir, usar una sierra, jugar tenis, sacar una muela, etc.; para enunciar los pasos del procedimiento es común preguntarnos ¿Qué comportamientos importantes muestran las personas que pueden ejecutarlo con éxito? Si lo que está sujeto a evaluación es un producto; como una pintura, un proyecto, un guión, un plano, una maqueta, una escultura, un plan de clase, etc.; para la enumeración de elementos a juzgar, conviene dar respuesta a la pregunta ¿Qué aspectos son característicos de la alta calidad de productos de este tipo? Finalmente si lo que es objeto de estudio es algún rasgo socio-personal o actitud como extroversión, veracidad, paciencia, cooperación, educación, generosidad, etc.; en este caso la pregunta pertinente es ¿Qué aspectos o acciones importantes se asocian comúnmente a esta actitud o rasgo socio-personal?

Una vez enumeradas las características o elementos a juzgar, en el caso de las escalas de evaluación, es muy importante la definición de la escala de medición para cada elemento o característica enunciada. Aunque una escala define un continuo de fondo, en realidad marcamos ciertos hitos que no necesariamente representan unidades significativas; es decir que no podemos asumir que todas tengan el mismo valor y que la distancia entre cada una de ellas sea igual, como sucede con escalas para medir alguna magnitud física. No obstante, en el momento de construir una escala podemos representar gráficamente distancias iguales entre los valores adoptados. Una recomendación útil para definir estos valores consiste en primero definir los extremos del continuo y enseguida los puntos intermedios, normalmente se usan escalas impares (3 a 5 niveles), siempre que se pueda detectar un punto neutro o un punto medio claro entre los dos extremos.

Los tres tipos más comunes de escalas son la numérica, la gráfica y la descriptiva; aunque a veces se usan combinaciones de ellas. La escala descriptiva implica una

descripción verbal más completa de los puntos a lo largo del continuo. Presentamos a continuación algunos ejemplos de escalas; las de la figura 4 son ejemplos de escalas numéricas y representan grados o frecuencias; mientras que en la figura 5, presentamos una escala gráfica y una descriptiva.

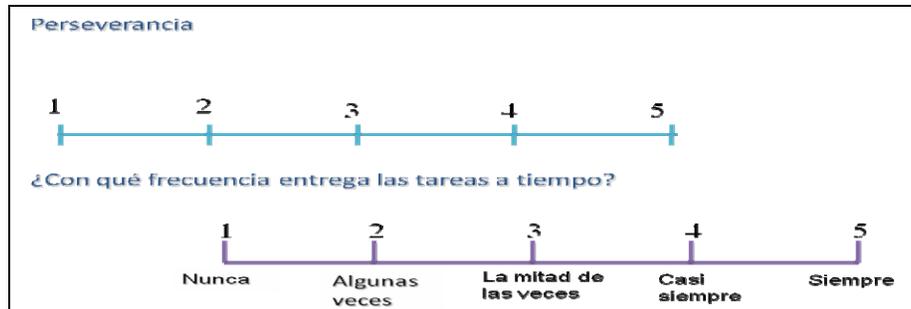


Figura 4. Escalas de evaluación numéricas

La escala descriptiva ha probado ser de gran utilidad para procesos de evaluación formativa ya que permite comunicar con mayor claridad a la persona evaluada acerca de su situación respecto a cada elemento de su ejecución o desempeño.

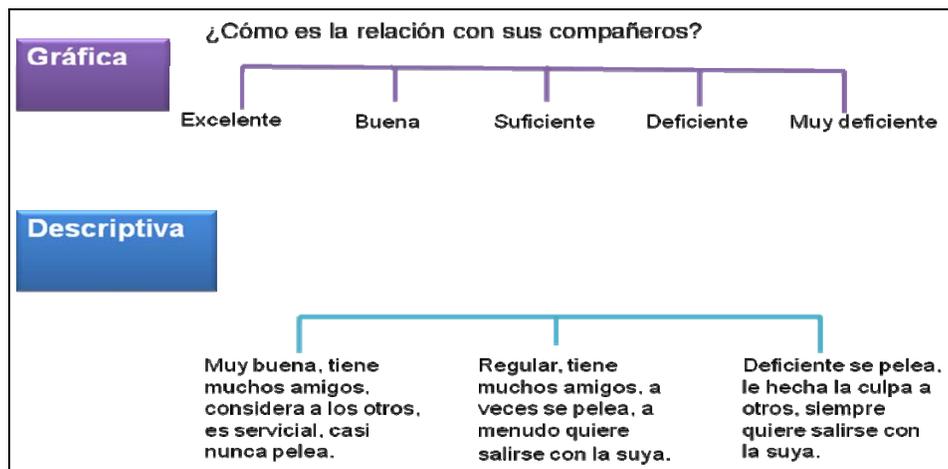


Figura 5. Escalas de evaluación gráfica y descriptiva

Para ser congruentes con los aspectos teóricos acerca de lo que significa la evaluación del aprendizaje, los procedimientos de evaluación deben contribuir al aprendizaje del estudiante y no sólo medirlo, adicionalmente el aprendizaje que se requiere para el desarrollo de competencias profesionales tanto genéricas como específicas a algún campo profesional de la educación superior es complejo y multidimensional por lo que requiere valorarse a través de múltiples formas (Murphy, 2006) aspecto que además requiere de estrategias de triangulación para el análisis de la información como se advierte en el paradigma propuesto por Casanova (2007) representado en la figura 2.

La evaluación de competencias

Hemos comentado ya, que la sociedad moderna delega en la educación superior la función de desarrollar en los estudiantes aquellas competencias que le permitan actuar de manera eficaz y responsable en dicha sociedad (Yániz y Villardón, 2006). La necesidad de ajustar los resultados pretendidos con lo evaluado requiere reflexionar acerca de la concepción de competencia como resultado de aprendizaje, para más información es conveniente revisar la concepción más actualizada de competencia en el marco de la educación superior que presento en (Leyva, 2009). En este espacio sólo comentaremos que:

En primer lugar la competencia supone **la adquisición de conocimientos, habilidades y actitudes**, por lo que es necesario recuperar evidencias de los tres tipos de adquisiciones.

En segundo lugar la competencia **supone la movilización estratégica** de estos elementos como recursos disponibles y necesarios para dar respuesta a una situación determinada en un contexto específico, por lo cual se requiere constatar la capacidad de movilizar los recursos de forma eficaz y ética para atender a una demanda.

En tercer lugar la competencia **se demuestra haciendo**, por lo que requerimos recoger evidencias de desempeño a partir de actividades planeadas que pongan en juego tales competencias.

En cuarto lugar el desarrollo de competencias **es un proceso de aprendizaje**, lo cual permite aprovechar las potencialidades de la evaluación para favorecer el logro de los objetivos formativos.

De acuerdo con Yániz y Villardón (2006) la evaluación en la formación universitaria cumple dos funciones fundamentales: la función sumativa de certificación de unos aprendizajes exigidos y la función formativa para favorecer el logro de dichos aprendizajes. Es decir que refiere a la evaluación en su función sumativa como evaluación de competencias y en su función formativa como evaluación para el desarrollo de competencias.

Esta concepción enlaza además con el concepto de **evaluación auténtica**, perspectiva que enfatiza el hecho de que una evaluación debe reflejar los desafíos, el trabajo y los estándares vigentes en el ejercicio profesional de manera que comprometa al alumno en forma interactiva con oportunidades para explicar, dialogar e investigar (Wiggins, 1989). Tal enfoque de evaluación requiere que el alumno actúe eficazmente con el conocimiento adquirido, en un amplio rango de tareas significativas para el desarrollo de competencias que permitan ensayar la realidad compleja de la vida social y profesional.

La competencia no puede ser observada directamente en toda su complejidad pero puede ser inferida a partir del desempeño, lo que lleva a pensar en los tipos de actuaciones que permitan reunir evidencia suficiente para hacer juicios razonables acerca de los logros de cada estudiante en relación con los resultados de aprendizaje pretendidos y claramente especificados, lo cual ayuda a los propios estudiantes a la comprensión de lo que se espera de ellos, a la vez que sirve para informar a otras personas interesadas sobre el significado de una calificación determinada (Yániz y Villardón, 2006) .

El portafolios de evaluación

Una de las herramientas en el contexto educativo que está experimentando un notable aumento en su utilización como sistema de evaluación ligado a la intervención educativa, es la práctica del portafolios de evaluación, lo cual representa un indicador de la búsqueda de mejoras en el contexto educativo y responde a la necesidad de subsanar algunas de las deficiencias cruciales derivadas de un modelo de enseñanza más estático y enciclopédico, que un conocimiento flexible de tipo condicional (Barberá, 2005).

El concepto del portafolios de evaluación existe desde hace tiempo en numerosos ámbitos; artistas, arquitectos, fotógrafos, diseñadores gráficos los usan para presentar sus trabajos a clientes potenciales. En educación son un fenómeno relativamente reciente y hasta ahora es que se empieza a explorar su potencial. A principios de los noventa se hablaba de su uso principalmente en referencia a la evaluación (Polin, 1991).

Un portafolio es un registro del aprendizaje que se concentra en el trabajo del alumno y en su reflexión sobre las tareas o actividades implicadas para su desarrollo; mediante un esfuerzo cooperativo entre el alumno y el docente, se reúne material indicativo del progreso hacia los resultados esenciales esperados (*National Education Association*, 1993). Usados en el marco de la evaluación, los portafolios son colecciones sistemáticas realizadas por alumnos y docentes que sirven como base para examinar los logros, los progresos, los procesos, las actitudes y el esfuerzo además de satisfacer las exigencias de responsabilidad habitualmente alcanzadas por procedimientos de prueba más formales (Johns, 1992) lo cual los hace pertinentes para enfoques mixtos.

Aunque existen diferencias de énfasis en las definiciones de portafolios, la mayoría comparten ciertas características esenciales:

1. Consisten en colecciones de la tarea de los alumnos, una serie de trabajos producidos por cada uno de ellos.
2. Las colecciones son deliberadas y no azarosas
3. Es una muestra de los trabajos reunidos con un propósito determinado.
4. Incluyen la oportunidad que tienen los alumnos de comentar su trabajo o reflexionar acerca de él.

Lo cierto es que el uso del portafolios de evaluación ha probado ser una herramienta muy valiosa de evaluación tanto sumativa como formativa, siempre y cuando se incorpore de forma planificada en el proceso de enseñanza y aprendizaje; su uso tiene muchas implicaciones psicopedagógicas ya que coadyuva tanto al fortalecimiento de las capacidades cognoscitivas de los alumnos, como a las metacognitivas, es decir de autorregulación de su proceso de aprendizaje. Esto es debido a que el portafolios promueve que el estudiante, mediante la documentación de evidencias de su aprendizaje, haga un seguimiento y reflexión de su propio desempeño, y con ello la evaluación continua a lo largo del tiempo, brindando información útil para la identificación de sus logros y fallas, con lo cual se facilita el ajuste progresivo de sus estrategias y hábitos de estudio.

De acuerdo con algunos investigadores y docentes (Agra, Gewerc y Montero, 2003; Seda-Santana, 2002; Barberá, 2005; Barragán, 2005; Arraiz y Sabirón, 2005), la estrategia de portafolio puede servir principalmente para involucrar a los alumnos con el contenido del aprendizaje, promover las habilidades de la reflexión y la autoevaluación, documentar su aprendizaje en áreas que no se prestan a la evaluación tradicional y facilitar la comunicación de los logros de los alumnos ante diferentes audiencias. Consecuentemente, el énfasis realizado en la práctica habitual del portafolios proporciona unos buenos hábitos cognitivos y sociales en los alumnos. En esta mayor participación en su propio proceso de aprendizaje, el alumno también aprende a responsabilizarse de la dirección que toma su educación, del nivel de profundización con el que decide tratar los contenidos y de los logros conseguidos.

Atendiendo a la necesidad que se tiene de contar con enfoques mixtos que por un lado favorezcan el valor estratégico de la evaluación formativa a la vez que se apoyen los juicios de acreditación académica que todo sistema educativo formal demanda como parte de su normativa, los portafolios han mostrado ser herramientas apropiadas en el cumplimiento de ambas funciones; además de que resulta idóneo para la evaluación de competencias, ya que permite integrar una gran variedad de tareas y actividades que pongan en juego los conocimientos, habilidades y actitudes de los estudiantes tanto en trabajo individual como de equipo.

Resulta especialmente útil para estos propósitos la propuesta documentada por Danielson y Leslie (1999), la cual consiste en tres etapas o portafolios: el de trabajo, el de evaluación y el de presentación. La primera fase o etapa denomina portafolios de trabajo contiene tanto trabajos en curso como muestras terminadas orientadas por competencias u objetivos de aprendizaje. Como su propósito es el de diagnosticar las necesidades de los estudiantes mediante evidencias de los procesos implicados en el desarrollo de las competencias u objetivos y en el nivel de logro alcanzado por cada alumno, permite cumplir la función formativa de la evaluación. En cuanto al proceso, se estructura en torno de un área de contenido específica o de una competencia y se documenta el progreso del alumno en el logro de las competencias u objetivos, mediante diversas evidencias de desempeño detectando lagunas o errores conceptuales o de procedimiento. El portafolio

de trabajo se revisa en su totalidad para evaluar sus elementos de manera continua. En esta parte, para cumplir adecuadamente la función formativa, resulta muy importante la retroalimentación del profesor para ayudar al alumno a detectar sus áreas de oportunidad.

Es posible trasladar algunos de los elementos del portafolios de trabajo a la siguiente fase o etapa la cual consiste propiamente en **el portafolio de evaluación**, el cual tiene la finalidad de documentar el cumplimiento de competencias u objetivos, en esta fase se cumple una función de carácter sumativo, es decir que se centra en resultados y permite la emisión de un juicio de acreditación académica.

También algunos de los elementos pueden trasladarse a un **portafolios de presentación** el cual tiene como propósito demostrar el nivel más alto de realización alcanzado por el alumno. Puede mantenerse a través de todo un ciclo agregando nuevos elementos cada año. El uso más gratificante para el alumno es el de presentar sus mejores trabajos e interpretar su significado, lo cual fortalece su motivación orientada a la tarea y sirve como un refuerzo de la autoeficacia.

Esta propuesta en fases, resalta la importancia de que sean los propios estudiantes quienes seleccionen lo que van a incorporar tanto al portafolios de evaluación como al de presentación, con lo cual ilustran lo que ellos creen que es importante de su aprendizaje, qué es lo que valoran y qué quieren mostrar a los demás. Estrategia que promueve el desarrollo de habilidades metacognitivas, es decir habilidades para la reflexión acerca de del propio pensamiento o conocimiento sobre nuestro aprendizaje, así como, sobre nosotros mismos en cuanto aprendices, lo cual se facilita si para ello tenemos evidencias tangibles de los avances y logros que se van documentando como parte del ejercicio de integrar los portafolios. Esta ventaja además de que resulta especialmente relevante ya que el estudiante se sensibiliza acerca de la importancia que tiene la documentación de todas las evidencias de su desempeño, es una fuente de motivación y de afirmación de la autoestima, aspectos que favorecen el logro de las competencias y desde luego la verificación de que se alcanzaron los estándares establecidos.

Hoy en día la autoevaluación y autogestión del conocimiento son características fundamentales que deben estar presentes en todo estudiante que pretenda iniciar estudios de educación superior o incorporarse al mundo del trabajo. La primera se refiere a los pensamientos personales de los alumnos en cuanto a sus conocimientos y habilidades, a sus estados afectivos en relación a su conocimiento, sus habilidades, motivación y características. La autogestión se refiere a su vez a la reflexión sobre el pensamiento en acción que ayuda al individuo a organizar los aspectos en la resolución de problemas (Klenowski, 2005).

En síntesis, la propuesta específica de esta guía práctica para la evaluación del aprendizaje, es que la estrategia del portafolios de evaluación se convierta en parte de la metodología de la enseñanza que se inserta en los procesos de la evaluación formativa de

los aprendizajes, de tal forma que sea posible la evaluación de trayectorias académicas y por lo tanto del desarrollo de competencias obtenidas por los alumnos en dichas trayectorias, lo que permite que existan más posibilidades de que los alumnos reciban retroalimentación de manera oportuna para apoyar su toma de decisiones acerca de su propio aprendizaje y que para los docentes constituya también una fuente de información útil para reorientar su práctica, logrando que su intervención sea acorde y oportuna para atender las necesidades y áreas de oportunidad detectadas por los alumnos.

Hemos hecho énfasis en la importancia que tiene asegurar la utilidad de la evaluación, lo cual se logra cuando se cumple la función formativa respecto a la retroalimentación como elemento fundamental que coadyuva a la mejora. Sin embargo esto sólo sucede si es efectiva y oportuna. Para ello es necesario tener claridad respecto a los estándares contra los cuales se va a contrastar el desempeño de los estudiantes para saber el grado en que cada uno de ellos se ha acercado a los estándares esperados en su ejecución o desempeño en las diversas actividades y tareas lo cual requiere como ya se comentó del trabajo colegiado en la fase de planeación para determinar los resultados de aprendizaje esperados. A partir de estos resultados esperados, se procede al análisis de las tareas o actividades que pongan en juego tales resultados; y en paralelo se trabaja en el establecimiento de los estándares de ejecución y desempeño que reflejen aquello que se espera que el estudiante demuestre durante la ejecución de las tareas y actividades.

La mejor forma de lograr tanto la orientación del alumno hacia el logro de tales estándares, así como la de proporcionar una retroalimentación efectiva y oportuna es la incorporación de rúbricas basadas en el establecimiento de estándares de ejecución y niveles de desempeño esperados en las diversas actividades y tareas. Las rúbricas se realizan una vez que se ha determinado la actividad o tarea a realizar para poner de manifiesto el desempeño del estudiante, lo que a su vez constituye una forma de garantizar que los profesores han logrado comunicar eficaz y oportunamente a los alumnos lo que se espera de ellos para cada actividad o tarea programada (Stevens & Levi, 2005). También servirán como referente para la detección oportuna de profesores y alumnos de aspectos logrados, así como de aspectos que requieren más trabajo para cerrar las brechas entre el estándar y lo que cada alumno es capaz de desarrollar en cada dimensión o criterio evaluado.

Los 4 componentes de las rúbricas de evaluación, organizados en términos del proceso para construirlas son:

1. Descripción de una tarea o actividad
2. Dimensiones: Son el tipo de habilidades que se ponen en juego para realizar la tarea, representan los criterios sobre los cuales vamos a centrar nuestra atención.
3. Escala de valoración: son los niveles de logro en la ejecución o desempeño
4. Descripción del nivel de ejecución de cada nivel de la escala (retroalimentación)

Debemos considerar estos componentes en la construcción de rúbricas, como se puede observar las rúbricas representan una escala de evaluación descriptiva por lo que podemos seguir los mismos pasos en su construcción. De manera similar a las listas de cotejo y las escalas de evaluación, se requieren dos insumos para que tenga sentido su construcción: Haber realizado la planeación en donde se especifiquen los resultados de desempeño esperados y haber elegido aquellas actividades o tareas que pongan en juego alguno o varios de estos resultados. Sólo así podremos definir las dimensiones o criterios implicados. Ahora bien, para definir una escala que sea apropiada, necesitamos antes determinar si el propósito de la evaluación es formativo, sumativo o mixto, generalmente se usan entre 3 y 5 niveles de ejecución o desempeño si es que vamos utilizar la técnica de observación de alguna ejecución o conducta en tiempo real; o de calidad en el caso en el que lo que estemos evaluando sea un producto.

Finalmente para poder describir cada uno del niveles de ejecución por cada dimensión o criterio, se requiere determinar el estándar o nivel de ejecución satisfactorio o de competencia máxima a lograr por cada dimensión. A partir de haber definido éste nivel se procede a describir, según sea el número de niveles de ejecución lo que representa cada uno de ellos. Facilita la tarea, pensar inicialmente en los extremos, por lo que después del nivel competente, se puede describir el nivel que implicaría que no se ha cubierto el estándar; es decir que la ejecución de la persona no es satisfactoria, y después pensar en él o los niveles intermedios y describir de manera cualitativa cada tipo de desempeño o ejecución. Es muy importante que las diferencias entre los niveles estén claramente delimitadas.

En términos de las principales ventajas que se han documentado del uso de las rúbricas podemos destacar las siguientes:

1. Proporcionan retroalimentación oportuna
2. La retroalimentación informa sobre el rendimiento en cada dimensión o componente de una tarea
3. Promueven el pensamiento crítico y la reflexión
4. Facilitan la comunicación con diversas audiencias
5. Nos permiten refinar los métodos de enseñanza

Como se puede deducir de estas ventajas las rúbricas son un aproximación adecuada cuando la función de la evaluación es formativa; no obstante, han probado ser una herramienta muy poderosa para la evaluación sumativa en el contexto de la educación basada en competencias, sobre todo en aspectos del aprendizaje que no pueden evaluarse a través de pruebas objetivas, porque requieren algún tipo de evidencia de desempeño. En todo caso para ambas funciones se requiere tener una metodología clara para la determinación de estándares de ejecución contra los cuales contrastar el desempeño de cada persona evaluada.

4. Elaboración de reactivos y diseño de pruebas objetivas para evaluación sumativa

La construcción de pruebas objetivas con reactivos de opción múltiple se recomienda principalmente para la evaluación sumativa, ya que nos permiten hacer un muestreo más representativo de los contenidos y procesos que se desea evaluar al final de un curso o de un período determinado.

Los estímulos de una prueba objetiva pueden ser de diferentes formas, sin embargo diversas investigaciones y la propia evidencia empírica resultante de las aplicaciones, demuestran que los reactivos de opción múltiple son los más estables y nos sirven no solo para medir los resultados de conocimiento, sino de una gran variedad de habilidades y capacidades intelectuales, por lo que resultan ser los de uso más común. Las investigaciones también han demostrado que los estímulos de opción múltiple con cuatro opciones son los mejores, siempre que se inicie con cinco opciones y se elimine la que consistentemente tenga menor frecuencia de respuestas.

Algunas ventajas de las pruebas objetivas conformadas por reactivos de opción múltiple son:

1. Se pueden adaptar fácilmente a los resultados de aprendizaje específicos.
2. Permiten un muestreo más adecuado del dominio a evaluar.
3. Se pueden calificar más rápido y objetivamente.
4. Son más sencillas de aplicar.
5. Poseen una sola respuesta correcta

Los reactivos de opción múltiple se conforman por una base y cuatro o cinco opciones de respuesta. La base es el enunciado que presenta la situación problema planteada explícita o implícitamente en una pregunta, afirmación o enunciado incompleto. Las opciones se entienden como las posibles respuestas, entre las cuales una responde correctamente al enunciado o pregunta y las otras son respuestas incorrectas llamadas distractores.

Sugerencias generales para elaborar un reactivo

En cuanto a su escritura:

- El lenguaje usado en la redacción de un problema debe ser apropiado para la materia que cubre, particularmente en lo que se refiere al vocabulario técnico.
- Debe redactarse con sencillez y corrección. El problema **no** debe medir habilidad para comprender estructuras gramaticales complejas, excepto cuando se está midiendo esta habilidad en particular.

- Evite errores gramaticales: de puntuación y de ortografía, así como abreviaturas.
- Evite el uso complicado de gerundios, participios, artículos y preposiciones.
- Emplee conceptos conocidos en lugar de sinónimos o vocabulario rebuscado.
- Si los artículos o preposiciones que acompañan a los sustantivos son iguales para todas las opciones, deben formar parte de la base.
- Tome en cuenta el tiempo de lectura por reactivo, utilice la información necesaria pero suficiente, para optimizar el tiempo.
- Evite elaborar reactivos que evalúen sólo el sentido común del sustentante.
- Tome en cuenta el nivel escolar y el de maduración de los examinados.

En cuanto al contenido:

- Tome como referencia la tabla de especificaciones.
- Elabore reactivos con base en los indicadores o resultados de aprendizaje de la tabla.
- Evite aumentar la dificultad eligiendo aspectos poco significativos del conocimiento.
- Incluya una sola idea al elaborar el reactivo.
- Evite evaluar conceptos de manera textual.
- Evite evaluar contenidos intrascendentes o triviales.
- No elabore preguntas capciosas.
- Los reactivos que evalúan comprensión siempre deben ser originales para evitar respuestas aprendidas de memoria.
- En la mayoría de los casos el problema debe contener sólo material relevante a su solución.

Indicaciones para elaborar la base

- Cada reactivo debe presentar, por sí mismo, un problema que al ser resuelto permita demostrar que se ha alcanzado un resultado de aprendizaje determinado.
- Incluya todos los elementos estrictamente necesarios para comprender el sentido correcto de la pregunta, de tal manera que sin leer las opciones tenga sentido propio.
- Los reactivos deben ser independientes unos de otros, sin que la información contenida en uno sugiera la solución de otro.
- Elabore un esquema de indagación y no sólo una palabra o frase.
- Cuide que la redacción sea precisa y adecuada.
- No sobrecargue el enunciado con material no pertinente por el deseo de seguir enseñando.
- Evite términos que confundan o den claves de la respuesta correcta.
- Evite que una pregunta ayude a responder otra.
- Redacte el enunciado en forma afirmativa siempre que sea posible.

Indicaciones para elaborar las opciones

- De las opciones (4) sólo una será correcta.

- La respuesta correcta debe resolver el problema satisfactoriamente.
- La respuesta correcta debe ser incuestionable (Por lo menos dos personas deberán revisar el reactivo para determinar si la solución correcta está clara)
- Incluya distractores plausibles.
- Deben tener concordancia gramatical con la base.
- Deben guardar entre sí un equilibrio coherente en su aspecto gramatical: sintaxis, género, número, tiempo verbal, etc.
- Deben ser congruentes con el campo semántico de la base.
- Evite repetir el material en cada opción e incluya el contenido común en la base.
- No debe repetirse ninguna opción ni con sinónimos.
- Deben presentar la misma extensión.
- Evite que una opción ayude a elegir la opción correcta, por indicios gramaticales o plausibilidad.
- Preséntelas al azar, a menos que se requiera un orden cronológico o numérico.
- Incluya en los distractores los errores más comunes de los estudiantes.
- No utilice en las opciones: *“todas la anteriores”, “ninguna de las anteriores”, “no sé”,* etc.
- Evite las formas negativas (*no, excepto*) o las absolutas (*nunca, siempre, completamente, totalmente*).
- Utilice enunciados verosímiles pero incorrectos. Evite las frases inadecuadas, ridículas o humorísticas.
- No repita palabras del enunciado entre las opciones.

Variantes con formato de opción múltiple:

En cuanto a variantes de reactivos de opción múltiple, podemos identificar al menos 5 tipos los cuales son más adecuados a cierto tipo de indicadores o resultados de aprendizaje a evaluar:

1. Simple
2. Jerarquización u ordenamiento
3. Selección de elementos de un listado
4. Relación de Columnas
5. Multireactivos (de cabecera o grupo de reactivos dependientes del contexto)

En el siguiente apartado brindamos una breve descripción de cada tipo, el formato general que presentan, así como ejemplos para facilitar la elaboración de reactivos que sean más acordes al tipo de indicador de aprendizaje a evaluar, sea este para evaluar conocimientos del dominio declarativo o procedimental. Generalmente para recabar evidencia de competencias, los multirreactivos presentan la posibilidad de utilizar una gran variedad de contextos y factores que permitan evaluar pensamiento de alto orden, necesarios para solución de casos y problemas y para la toma de decisiones sustentada.

1. Formato Simple:

Es el planteamiento de un problema en forma sencilla, sin embargo existen diversas formas de formular la base, por ejemplo con forma interrogativa, imperativa, de completamiento y de manera excepcional negativa.

Formato:

BASE: Texto y / ó gráficas. Si el reactivo contiene esquemas, gráficas o imágenes, es importante que incluya también las instrucciones para que se entienda con claridad lo que se tiene que hacer.

OPCIONES: (Textos y/ó gráficas)

- A)-----
- B)-----
- C)-----
- D)-----

Ejemplos:

¿En qué palabra está subrayada la sílaba tónica?

Interrogativa

- A) Avenida
- B) Consecuencia
- C) Revistero
- D) Consulado

Elija la opción que presenta una palabra grave.

Imperativa

- A) Paz
- B) Célula
- C) Fotografía
- D) Situación

La opción que presenta una preposición es:

Completamiento

- A) el
- B) en
- C) ha
- D) su

Las siguientes palabras son agudas EXCEPTO:

Completamiento Negativa

- A) situación
- B) potenciar
- C) consulado
- D) además

2. Jerarquización u ordenamiento

Este tipo de pregunta requiere elegir la opción que contiene el ordenamiento o secuencia correcta de diversos elementos. La base del reactivo indica que la situación por resolver es ordenar, con base en algún criterio, una serie de elementos presentados al azar en una lista. Las opciones se forman a partir de la combinación de varios ordenamientos y sólo una combinación es correcta.

Este tipo de reactivo es útil para evaluar entre otros aspectos: Identificación de procesos, procedimientos y eventos históricos. El número de elementos a ordenar no deber ser muy grande y se asignará un punto por el ordenamiento correcto y ninguno por los demás.

Formato:

ENUNCIADO: -----.	
1. -----	} Aquí se escribe la lista de elementos en desorden -
2. -----	
3. -----	
4. -----	
5. -----	

OPCIONES:

- A) 3, 2, 4, 1, 5
- B) 2, 4, 3, 5, 1
- C) 5, 1, 2, 4, 3
- D) 4, 3, 1, 2, 5
- E) 2, 3, 4, 5, 1

Ejemplo:

Indica la secuencia en la que se ordenan cronológicamente los acontecimientos de la vida política de Benito Juárez.
1. Derrocó a Santa Anna
2. Se convirtió en gobernador de Oaxaca
3. Fue desterrado a Cuba y Nueva Orleans
4. Contribuyó a la caída de Maximiliano I

- A) 2, 3, 1, 4
- B) 1, 2, 4, 3
- C) 2, 4, 3, 1
- D) 4, 1, 3, 2

3. Selección de elementos de un listado:

Es el formato de reactivo mediante el cual se le solicita al sustentante que elija de entre distintos elementos, una serie de ellos menor a la enunciada. Este tipo de formato permite evaluar la capacidad de discriminación de características, datos, elementos, circunstancias, eventos y reglas, entre otros aspectos; de un proceso, fenómeno o procedimiento determinado.

Formato:

ENUNCIADO: -----.	
1. -----	} Aquí se escribe la lista de elementos mayor a la solicitada en la base.
2. -----	
3. -----	
4. -----	
5. -----	

OPCIONES:

- A) 1, 2, 5
- B) 2, 3, 5
- C) 1, 2, 4
- D) 2, 4, 5



Se elige un número de elementos relacionados con un aspecto específico

Ejemplo:

De los siguientes tipos de tejido presentados en la lista, elige los cinco tipos de tejido conectivo.
1. Muscular estriado
2. Óseo
3. Cartilaginoso
4. Conjuntivo denso
5. Nervioso
6. Adiposo
7. Sanguíneo

- A) 1, 3, 4, 5, 7
- B) 2, 3, 4, 6, 7
- C) 1, 2, 3, 6, 7
- D) 2, 3, 4, 5, 6

4. Formato de relación de columnas:

Es una forma modificada del estímulo de opción múltiple donde se enlistan en una columna una serie de *premisas* y en la otra columna las *respuestas* posibles. No obstante, a menos que todas las respuestas sirvan como opciones razonables para cada premisa, el formato será inadecuado. Hay dos errores comunes al elaborar este tipo de reactivos:

- Tendencia a mantener la misma cantidad de premisas y respuestas;
- Incluir demasiados elementos en ambas listas.

Reglas para su elaboración

- Incluya sólo material homogéneo en cada estímulo,
- Utilice listas de tamaño razonable y ubique las respuestas a la derecha; esto permite una mayor facilidad de lectura y búsqueda.
- Utilice un número de respuestas menor o mayor que el de las premisas y haga que las respuestas se puedan usar más de una vez, para reducir la respuesta por tanteo.
- Especifique en las instrucciones las bases de la correspondencia e indique que cada respuesta puede usarse una o más veces, o no usarse.

Formato:

ENUNCIADO: (Instrucciones y descripción en la base-----).	
<p>TÍTULO DE LA COLUMNA IZQUIERDA</p> <p>1. _____</p> <p>2. _____</p>	<p>TÍTULO DE LA COLUMNA DERECHA</p> <p>a) _____</p> <p>b) _____</p> <p>c) _____</p> <p>d) _____</p>

OPCIONES:

- 1 a, b; 2 c, d
- 1 a, c; 2 b, d
- 1 b, c; 2 a, d
- 1 c, d; 2 a, b

Ejemplo:

Relacione los movimientos con sus consecuencias.	
Movimientos	Consecuencias
<p>1. Rotación</p> <p>2. Traslación</p>	<p>a) Cambio aparente del tamaño del sol</p> <p>b) Desviación de los cuerpos en caída libre hacia el Este</p> <p>c) Deformaciones de la Tierra</p> <p>d) Estaciones del año</p>

- 1 a, b; 2 c, d
- 1 a, c; 2 b, d
- 1 b, c; 2 a, d
- 1 c, d; 2 a, b

5. Multirreactivos

Es el formato mediante el cual podemos evaluar de forma integrada, una variedad de conocimientos y habilidades cognoscitivas. Pueden ser de base teórica y metodológica y de distinto nivel taxonómico. Este formato se recomienda para plantear un problema o describir un contexto y con base en él elaborar una serie de reactivos. Su construcción debe presentar consistencia y claridad en el problema presentado en la base y en las preguntas que se desprenden de él.

Formato:

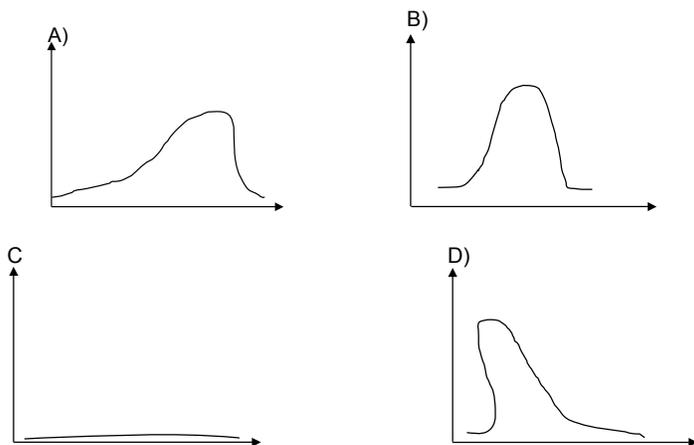
Instrucciones:----- Situación o caso : -----.
Base 1 1. -----
Opciones: A) ----- B) ----- C) ----- D) -----
Base 2 2. -----
Opciones: A) ----- B) ----- C) ----- D) -----
Se incluyen todos los reactivos que se estime conveniente

Ejemplo 1. Estadística para la carrera de Psicología

Instrucciones: Lea el siguiente caso y responda las siguientes cuatro preguntas:

Se aplicó un inventario de habilidades de estudio, en su versión autoreporte a 900 alumnos de las licenciaturas de psicología, ingeniería y derecho. Los resultados indican que los alumnos de psicología presentan una distribución negativamente sesgada en la escala de autorregulación; los alumnos de ingeniería muestran un sesgo positivo en la escala de adquisición de la información y; los alumnos de derecho muestran distribuciones con sesgo casi igual a cero en las escalas de autorregulación y adquisición de información.

1. Forma de distribución que representa mejor las puntuaciones de los alumnos de Psicología



Razonamiento.

La respuesta correcta es A), ya que una curva con sesgo negativo extiende una de sus colas hacia los puntajes bajos y acumula la mayoría de los datos en las puntuaciones altas. La opción C) es incorrecta por que hace referencia a una curva aplanada o con baja curtosis, D) no es correcta por que hace referencia a curvas con sesgos positivos y la opción B es incorrecta por que hace alusión a una curva con sesgo casi igual a cero.

2. ¿De acuerdo con la distribución de frecuencias, ¿cómo es la mayoría de los alumnos de ingeniería?

- A) Una proporción mayor de alumnos altamente estratégicos en adquisición de información
- B) Todos los alumnos son medianamente estratégicos en adquisición de información
- C) Una proporción mayor de alumnos bajos en estrategias de adquisición de información
- D) Todos los alumnos son altamente estratégicos en adquisición de información

Razonamiento.

La respuesta correcta es C) ya que una curva con sesgo positivo extiende una de sus colas hacia la derecha y acumula la mayoría de los datos en las puntuaciones bajas. A) es incorrecta porque representa una curva con sesgo negativo. B) es incorrecta porque representa una curva con sesgo igual a cero. D) es incorrecta porque hace referencia a una curva aplanada o con baja curtosis.

3. ¿Cómo son los valores de las medidas de tendencia central de las puntuaciones de los alumnos de derecho en cuanto a sus estrategias de adquisición de la información?

- A) la moda es mayor que la media y la mediana
- B) la moda es menor que la media y la mediana
- C) la media es diferente de la moda y la mediana
- D) la media es igual que la moda y la mediana

Razonamiento.

La opción correcta es D) ya que en una curva con sesgo casi igual a cero; la media, la mediana y la moda se ubican en el mismo valor debajo de la curva. Las opciones A), B) y C) son incorrectas por que se refieren a valores tomados de curvas con sesgo positivo y negativo.

4. ¿Cómo son la mayoría de los alumnos de derecho en sus estrategias de adquisición de información, en comparación con los alumnos de ingeniería?

- A) con mejores estrategias de adquisición de información que los ingenieros
- B) casi igual de estratégicos en adquisición de información de los ingenieros
- C) menos estratégicos en adquisición de información que los ingenieros
- D) muy por debajo de los ingenieros en adquisición de información

Razonamiento.

La opción correcta es A) ya que los alumnos de derecho al mostrar una distribución con sesgo casi igual a cero poseen una distribución casi normal, mientras que los ingenieros tienen una distribución con sesgo positivo. Las opciones A), B) y D) son incorrectas porque se refieren a situaciones en las cuales los alumnos de derecho presentarían curvas con sesgo positivo o negativo.

Ejemplo 2. Carrera de Odontología: Área de formación operativa

Lee con atención la información del siguiente caso clínico y contesta las dos preguntas asociadas:

Se presenta a consulta un paciente de 12 años de edad. Durante la exploración se observa incisivo central izquierdo de color pardo y el incisivo central derecho de coloración rojiza, ambos incisivos se muestran sintomáticos; en la radiografía se observa tratamiento de conductos con muy poca estructura radicular en el central izquierdo y en el incisivo central derecho una zona radio lúcida en forma de bombilla a nivel del tercio medio del conducto. Al interrogarlo, el paciente refiere haber sufrido fractura de ambos incisivos superiores centrales a la edad de siete años, con exposición de la pulpa vital en el central izquierdo, al cual se le realizó el tratamiento de conductos de inmediato.

1. ¿Qué procedimiento se debió haber realizado para preservar más estructura radicular en el central izquierdo?

- A) Tratamientos de conductos
- B) Apicogénesis
- C) Observación y control
- D) Implante endodóntico

Respuesta correcta: B)

Razonamiento.

A la edad de siete años el ápice los incisivos centrales no se encontraría totalmente formado, por lo que el tratamiento inicial en caso del central izquierdo, con la pulpa dental expuesta debió ser la apicogénesis o pulpotomía con hidróxido de calcio para permitir la formación radicular y el cierre apical, y posteriormente –ya finalizado éste- efectuar el tratamiento de conductos.

2. ¿Qué tratamiento se recomienda para el incisivo central derecho?

- A) Observación y control
- B) Apicogénesis
- C) Cirugía exploratoria

D) Tratamientos de conductos

Respuesta correcta D)

Razonamiento.

En el caso del central derecho el diagnóstico es muy claro, dado que se describen las características clínicas y radiográficas que corresponden a una resorción radicular interna, en cuyo caso se requiere realizar el tratamiento de conductos para evitar que ésta continúe.

Ejemplo de formato para captura de reactivos.

No.	Concepto	Estándar o especificación
0	Identificación	Nombre del redactor, fecha y número de reactivo
1	Rasgo del perfil a evaluar e indicador.	Incluya aquí el rasgo del perfil referencial de validez correspondiente. En seguida el indicador específico que evalúa el reactivo que va a redactar.
2	Clasificación por dominio y por nivel taxonómico	Clasifique los reactivos de acuerdo con el dominio, es decir tipo de conocimiento: Información, procedimiento mental, procedimiento psicomotor. Clasifique los reactivos de acuerdo con los niveles de taxonómicos de diseño Nivel 1. Recuperación Nivel 2. Comprensión Nivel 3. Análisis Nivel 4. Utilización del conocimiento
3	Redacción de la base del reactivo.	Siga los lineamientos de redacción presentados en el punto 3.3 Recomendaciones Generales, de este manual. Redacte el enunciado de forma clara, no use formas negativas, no utilice determinantes específicos (siempre, nunca, etc.). Se recomienda seleccionar el tipo de formato (simple, elección de un listado, relación de columnas, jerarquización y multirreactivo) más apropiado al tipo de resultado de aprendizaje a evaluar. Se pueden incluir dibujos, imágenes, cuadros, etc., dependiendo de lo que se desea evaluar.
4	Respuesta correcta y distractores	Todos los reactivos deben ser de 4 opciones. Coloque la respuesta correcta en la opción A - para fines del Banco de Reactivos si las opciones pueden presentarse aleatoriamente. Excepto si el reactivo tiene las opciones "fijas", en cuyo caso debe indicarse cuidadosamente la respuesta correcta. El reactivo debe incluir una respuesta correcta, única y demostrable.
5	Razonamiento	Es muy importante incluir la justificación, procedimiento o explicación para llegar a la respuesta correcta. Se explicará también cómo se construyeron las respuestas incorrectas o distractores.
6	Referencias	El reactivo debe incluir por lo menos una referencia de donde se haya obtenido la información, con este formato: Autor (fecha) "título", editorial, país, páginas.

Análisis y calibración de reactivos

En términos generales el procedimiento de calibración de reactivos sirve para el diseño de los exámenes y consiste en los siguientes pasos:

- Elección de una muestra representativa de sustentantes para la aplicación piloto de los exámenes.
- Integración de una versión de cada examen para su aplicación en el estudio piloto.
- Análisis de reactivos.
- Los reactivos que en la calibración no cumplan con los parámetros mínimos de dificultad y discriminación recibirán una primera revisión por parte del supervisor, quien estimará si el reactivo merece ser rediseñado o desechado definitivamente. En caso de rediseño, el reactivo será revisado por el redactor del reactivo, por el revisor y por el supervisor, con objeto de obtener una nueva versión.
- Los reactivos que cumplan con los parámetros requeridos, serán considerados como "aprobados" y se darán de alta definitiva en el Banco de reactivos del examen correspondiente, con los índices psicométricos obtenido en el piloteo.
- Los reactivos clasificados como "desechables" serán eliminados del Banco.

Calibración de reactivos

Toda pregunta en una prueba tiene un componente de varianza de error adjudicable al evaluador y no al sustentante por esto debemos verificar si la pregunta es pertinente al grupo de estudiantes al que está siendo aplicada.

Para las pruebas de aprovechamiento utilizadas para evaluación en aula existen procedimientos sencillos que nos permiten asegurar una mayor calidad en nuestros procesos de evaluación. Sólo se justifican procedimientos más complejos cuando se requiere una mayor precisión en procesos de evaluación a gran escala.

Las preguntas tienen una dificultad relativa de acuerdo con el grupo y nivel en que se apliquen. Una pregunta con una baja frecuencia de aciertos tiene una dificultad mayor que otra con una frecuencia alta. La dificultad del reactivo es una información básica para la base de datos del banco de reactivos y para el momento del diseño de la prueba, ya que ésta debe estar balanceada en cuanto a la dificultad global y de cada una de sus áreas o secciones.

El índice de dificultad de acuerdo con el modelo clásico, será la frecuencia de respuestas correctas entre el número total de alumnos y se expresa en términos de porcentaje:

$$GD = \frac{P}{N} \times 100 \quad ; \quad 0 \leq GD \leq 100\%$$

donde

P = Total de respuestas correctas de los sustentantes

N = Total de sustentantes (sólo respuestas válidas)

Los parámetros generalmente aceptados por los evaluadores para el índice de dificultad clásica es:

$$20\% \leq GD \leq 80\%$$

Otro indicador importante a obtener es el poder de discriminación, es decir la capacidad del reactivo para distinguir entre sustentantes de alto y bajo rendimiento ya que nos permite determinar si la pregunta la aciertan más frecuentemente los estudiantes del grupo superior, comparados con los estudiantes del grupo inferior. Si la prueba y las preguntas están bien elaboradas, se espera que el estudiante que conteste mayor cantidad de preguntas corresponda al que obtiene un mayor rendimiento en la prueba y viceversa.

Mientras que la dificultad es el porcentaje de respuestas correctas al reactivo entre el total de personas evaluadas, el poder de discriminación es la diferencia de respuestas correctas entre el grupo superior y el grupo inferior y se calcula mediante la fórmula:

$$PD = \frac{P(GS) - P(GI)}{N} \times 100$$

donde:

P(GS) = Respuestas correctas del Grupo Superior

P(GI) = Respuestas correctas del Grupo Inferior

N = Total de sustentantes (sólo respuestas válidas)

Desde luego se espera que el índice de discriminación sea mayor a cero, de lo contrario el reactivo se rechaza. El criterio de aceptación de un reactivo en términos de su poder de discriminación debe ser estricto, por lo que conviene que el poder de discriminación sea mayor o igual al 30% del grado de dificultad del reactivo.

Para facilitar la interpretación del poder de discriminación podemos usar la relación discriminativa la cual está incluida en el reporte del análisis de reactivos, para que un reactivo sea aprobado se requiere que esta relación sea mayor o igual a uno ($RD \geq 1$). La fórmula es:

$$RD = \frac{PD}{ND}$$

Los reactivos con valores inferiores de RD deben revisarse personalmente por los redactores de reactivos para verificar su redacción, la opción correcta, el comportamiento de las opciones distractoras, etc.

Estos indicadores clásicos del reactivo se pueden obtener de manera automática con Kalt de uso general. También se puede calcular mediante otros procedimientos específicos de teoría de respuesta al ítem, si se obtiene el paquete de Kalt Criterial.

Análisis de distractores

Para analizar cada una de las opciones de respuesta, se construyen tablas de contingencias con las opciones de respuesta y la frecuencia de respuestas de los sustentantes divididos en Grupos Superior (GS) e Inferior (GI) como una partición de la población sustentante, dividida en términos de la mediana.

Para que un reactivo sea aceptable se deben verificar varias relaciones. Las verificaciones a realizar corresponden al funcionamiento de un reactivo en la tabla de contingencias, de acuerdo con las siguientes condiciones:

- a) Se esperan más respuestas correctas en el Grupo Superior que en el Inferior.
- b) Se espera que la diferencia en respuestas del Grupo Superior e Inferior sea por lo menos igual a un valor normativo de 0.3 del total de respuestas correctas.
- c) Se espera que los distractores sean más atractivos para el Grupo Inferior que para el Superior.
- d) Se espera que los distractores sean igualmente plausibles, es decir que no haya distractores más atractivos que otros y que no haya uno en particular que compita con la respuesta correcta.

Todas estas comparaciones fijan la validez de las opciones en términos de la dificultad del reactivo y del azar de las respuestas de los sustentantes. Los reactivos que no cumplen con el patrón de cotejo por dificultad y por azar deben ser revisados o eliminados. Estas relaciones establecen, asimismo, los conjuntos de valores esperados por azar o por grado de dificultad, con los cuales se pueden efectuar pruebas de hipótesis χ^2 (Chi cuadrada) para verificar la diferencia entre valores observados y valores esperados. Esta información ayuda a los redactores en la revisión y mejoramiento de aquellos distractores que no funcionan adecuadamente. El procedimiento está incluido en el software y permite el análisis de todas y cada una de las preguntas de la prueba para mejorar la calidad de la evaluación.

Para evaluar la dificultad del reactivo de una manera rápida y accesible utilizaremos la medida clásica del grado de dificultad como porcentaje de respuestas correctas. También se puede calcular mediante otros procedimientos como en el caso del

análisis de Rasch. En el Manual Técnico del CENEVAL (1999), se describen ambos procedimientos.

Ejemplo para un cálculo rápido:

Supongamos que tenemos un grupo de 40 estudiantes y les aplicamos una prueba objetiva con reactivos de cinco opciones de respuesta. Después de calificar a los alumnos, los ordenamos desde el que obtuvo el mayor porcentaje de respuestas correctas hasta el de menor porcentaje, y dividimos al grupo por la mediana de manera que nos quedan dos grupos iguales a los cuales designaremos como grupo superior y grupo inferior.

Opciones	A	B*	C	D	E
Superior 20	1	12	4	2	1
Inferior 20	5	3	3	7	2
Total 40	6	15	7	9	3

* = respuesta correcta

El índice de dificultad de acuerdo con el modelo clásico, será la frecuencia de respuestas correctas (15), entre el número total de alumnos (40) y se expresa en términos de porcentaje:

$$GD = \frac{P}{N} \times 100 \quad ; \quad 0 \leq GD \leq 100\%$$

donde

P = Total de respuestas correctas de los sustentantes

N = Total de sustentantes (sólo respuestas válidas)

En nuestro ejemplo:

$$GD = 15/40 = 0.375 * 100 = 37.5$$

El reactivo del ejemplo tiene un valor aceptable porque se encuentra entre el 20% y el 80%, (aún con el criterio más estricto de 27% y 73% que exige Kalt), observamos además que se trata de un reactivo difícil ya que lo contestaron menos del 50% de los alumnos.

En nuestro ejemplo, el poder de discriminación del reactivo es:

$$\begin{aligned} PD &= (12 - 3) / 40 \\ &= 9 / 40 \\ &= 0.225 \\ &\times 100 = 22.50 \end{aligned}$$

Y la norma discriminativa es:

$$\begin{aligned} ND &= 0.3 \times GD \\ \text{es decir:} & \quad 30\% \text{ del GD} \\ & \quad 30\% (37.5) = 11.25 \end{aligned}$$

La relación discriminativa:

$$\begin{aligned}RD &= 22.5 / 11.25 = 2 \\ 2 &\geq 1\end{aligned}$$

por lo que el reactivo que estamos analizando es aceptable.

Análisis de la prueba

Desde la planeación hemos trabajado en garantizar la validez de los exámenes que estamos construyendo, no obstante, además de la validación lógica de contenido de cada reactivo de la prueba, es necesario hacer algunos análisis empíricos a partir de resultados de las aplicaciones, que nos permitan verificar la calidad de la prueba.

La tabla de especificaciones y el perfil referencial de validez elaborados por grupos colegiados o comités académicos, constituyen los fundamentos de la validez teórica del instrumento. Son también instrumentos base para los procedimientos de validación por jueces, lo cuales tienen que emitir juicios respecto a si los reactivos son adecuados y pertinentes al perfil referencial y a la tabla de especificaciones, y si el muestreo del universo es representativo. Al final de este documento se proporciona un ejemplo de formato que han de llenar los jueces que validen cada reactivo de una prueba objetiva (Anexo).

El concepto de validez

Desde un punto de vista científico, la validez de constructo es en sí la validez, mientras que los demás tipos de validez (predictiva, concurrente, convergente, factorial) pueden ser considerados más bien como estrategias de colección y análisis de datos empleadas para probar las conexiones conceptuales entre la medición y el constructo (Angoff, 1988; APA, AERA & NMCE, 1999). Aquí el énfasis está dado en sustentar el grado en que los puntajes en la prueba representan la medida de la característica o atributo psicológico que se supone evalúa la prueba, es decir, el constructo teórico, por lo que es necesario establecer procedimientos para la revisión lógica de tal adecuación, el análisis de su estructura interna o el análisis de la relación de la prueba con variables externas (caso de la validez concurrente, o predictiva).

La validez de una inferencia depende de la intención específica de la prueba, del procedimiento usado, así como de las condiciones de la colección de los datos, de los procedimientos de enjuiciamiento y medición seguidos, del análisis de los procedimientos empleados y de las características de las personas que proporcionan los datos y la información. Estas dependencias originan la identificación de diversas fuentes de evidencia que pueden aclarar diversos aspectos de la validez, lo cual no significa que sean

distintos tipos de validez. La validez es un concepto unitario y se refiere al grado en el cual la evidencia acumulada apoya la interpretación de las medidas de la prueba de acuerdo con el propósito propuesto.

En función de estas formas diversas de obtener evidencias de validez, se adoptaron categorías tales como la validez de contenido, de constructo y de criterio, esta última subdividida a su vez en dos categorías denominadas predictiva y concurrente. Esto no implica que sean diferentes tipos de validez o que alguna estrategia de validación sea mejor para cada tipo determinado de inferencia o uso posible de una prueba; de hecho no es posible hacer una distinción rigurosa entre estas categorías.

En la nueva versión de estándares de la APA, AERA, NMCE (1999) no se sigue ya con esta clasificación; en cambio, se describen las fuentes de evidencia que pueden emplearse para evaluar una interpretación propuesta de las medidas de una prueba para un propósito particular. En esta publicación de estándares se presentan cinco fuentes o líneas de evidencia de validez, las cuales se describen en una breve introducción del capítulo que agrupa los 24 estándares de validez. En el glosario del documento, los autores proporcionan definiciones de los términos, explicando la diferencia entre el uso tradicional y el actual.

Análisis de confiabilidad

Además de la validez, debemos demostrar que los exámenes que construimos son confiables, es decir que podemos obtener la misma medición en diferentes condiciones. Confiabilidad indica consistencia y precisión en la medición.

En algún caso la condición será el paso del tiempo (*estabilidad temporal*), es decir, que sin que medie ningún aprendizaje extra, la persona que es medida en dos momentos diferentes debe dar las mismas respuestas al instrumento. Por el dinamismo que representa el proceso de aprendizaje el primer caso está expuesto a cambios en cortas temporadas, midiéndose en muchas ocasiones la *estabilidad* del instrumento viciada con el dinamismo de aprendizaje, por lo que no se sugiere realizar estudios de este tipo en períodos que excedan los tres meses.

En un segundo caso la condición será similitud de instrumentos (*equivalencia*), es decir, que un estudiante deberá obtener calificaciones similares si se le aplican dos instrumentos de medición formados con la misma cantidad de preguntas seleccionadas con índices de dificultad similares con base en una tabla de especificaciones común.

Las mediciones en el área educativa, a diferencia de aquellas efectuadas en las ciencias naturales, difícilmente mantienen una precisión exenta de variabilidad entre un instrumento y otro a pesar de los grandes esfuerzos por buscar la equivalencia.

En el tercer caso la condición consiste en la relación que guardan cada una de las partes del instrumento, con el resultado final que se obtiene (*consistencia interna*). Esto implica la relación que guarda la calificación total del estudiante con las calificaciones obtenidas en los temas y las preguntas que integran el instrumento.

La consistencia interna es, en la mayoría de los casos, la forma más estable de medir la confiabilidad de un instrumento. Además, un estudio de consistencia interna permite la obtención automática de la validez teórica, siempre y cuando el universo de conocimientos esté bien definido y el muestreo de los conocimientos sea representativo. Por esto y, asumiendo el cuidado requerido en la elaboración de la tabla de especificaciones y la elaboración de las preguntas que representen el conocimiento en forma amplia y en la taxonomía requerida, se propone una forma breve de obtención de la consistencia en el instrumento, incluyendo ambos índices requeridos.

La confiabilidad de un instrumento se acostumbra estimar por medio de la Alfa de Cronbach (o de Cronbach-Hoyt), la cual parte de un análisis de varianza múltiple, bajo ciertas hipótesis simplificadoras. En el software Kalt, se obtiene el valor del Alfa mediante una forma directa, empleando solamente tres variables: el número de personas, la varianza muestral y los grados de dificultad de los reactivos, que son los parámetros empleados generalmente en nuestro medio (Tristán, 1996).

La fórmula directa para el cálculo de la confiabilidad es:

$$\text{ALFA} = \frac{N}{N-1} \left(1 - \frac{\sum GD(100 - GD)}{s^2 N^2} \right)$$

donde

N = número de reactivos

GD = Grado de dificultad de cada reactivo

s^2 = varianza de puntuaciones de los sustentantes en porcentaje de aciertos

En cada aplicación se obtiene de manera rutinaria la confiabilidad de todo el instrumento, así como por tema o área del examen. Los valores aceptables de ALFA se cotejan contra un modelo en términos del número de reactivos, buscando en todos los casos los valores más altos de ALFA.

Análisis de error de medida.

Mediante el software Kalt, se obtiene en cada corrida en forma rutinaria diversos errores de medida (Tristán, 1999). En particular se tienen los siguientes errores estándar:

a) error estándar para la media = $\frac{\sigma}{\sqrt{N}}$

b) error estándar para la mediana = $\sigma\sqrt{\frac{\pi}{2N}}$

c) error estándar para la desviación estándar = $\frac{\sigma}{\sqrt{2N}}$

d) error estándar de medida = $s\sqrt{1-\alpha}$

e) error de medida de cada sujeto = $t_{0.975}(M-1)\frac{s}{\sqrt{M-1}}$

en estas expresiones:

M = número de reactivos

N = número de sustentantes

s = desviación estándar de las dificultades de los reactivos

$t_{0.975}(M-1)$ = valor de t de Student para 97.5% de confianza (5% en prueba estadística de 2 colas) para M reactivos (M-1 grados de libertad)

α = confiabilidad (ALFA de Cronbach)

σ = desviación de las puntuaciones

π = 3.14159265359

Longitud del examen y de las secciones

Tanto los valores de confiabilidad como de error de medida, obtenidos de manera rutinaria, nos permitirán apoyar la toma de decisiones respecto a la determinación de la longitud más adecuada de la prueba y de las secciones.

La longitud del instrumento se determina en función de estos elementos básicos: las proporciones de reactivos especificadas en la tabla de validez de contenido, el tiempo disponible para la aplicación del cuestionario y el error de medida deseable para dictaminar el nivel de desempeño o dominio de un sustentante. En relación con las proporciones y el tiempo, los consejos técnicos deciden en función de experiencias previas o de opiniones sometidas a consenso. En el caso del error, en cambio, se dispone de un modelo matemático.

A partir de un modelo procedente de la teoría de muestras pequeñas se pueden encontrar límites inferiores para el valor del error de medida que conducen, a su vez, a

longitudes de un examen o partes del mismo (Tristán, 1999). El tamaño de la muestra de reactivos que debe aplicarse a los sustentantes tiene un error dado por la expresión:

$$\text{error de medida} \geq t_{0.975}(M - 1) \frac{s}{\sqrt{M - 1}}$$

donde:

M = número de reactivos

s = desviación estándar de las dificultades de los reactivos

$t_{0.975}(M-1)$ = valor de t de Student para 97.5% de confianza (5% en prueba estadística de 2 colas) para M reactivos (M-1 grados de libertad)

Se puede obtener una tabla en donde se presenten algunos valores indicativos de la longitud del examen o de sus partes, en relación con el error de medida esperado:

error %	20	15	12	10	5
Número mínimo de reactivos	12	19	27	39	131

La información estadística que se obtiene a partir de la teoría clásica de análisis de los resultados de una aplicación de los instrumentos, es útil para el diseño de versiones de exámenes válidos, confiables y objetivos, además de que permite la mejora continua tanto de reactivos, como de las versiones de examen que se diseñen a partir de los mismos, razón por la cual siempre se recomienda hacer una aplicación piloto de los exámenes, es decir una aplicación que en la medida de lo posible replique las condiciones de una situación de prueba, principalmente en términos de la representatividad de la población a la que se dirige.

En el caso de la aplicación de exámenes objetivos a gran escala y de alto impacto; los estándares de calidad de la APA (1999) exigen realizar análisis de reactivos y del comportamiento de la prueba; en estos casos resulta más apropiado contar con modelos logísticos de análisis para integrar bancos de reactivos a partir de los cuales se puedan generar exámenes que cumplan tanto las especificaciones de una tabla de validez de contenido, como los estándares de calidad necesarios para garantizar un proceso evaluativo técnicamente impecable.

Referencias Bibliográficas

- Agra, M. J.; Gewerc, A. y Montero, L. (2003): El portafolios como herramienta de análisis en experiencias de formación on line y presenciales. España, *Enseñanza*, 21.
- American Psychological Association, American Educational Research Association, National Council on Measurement in Education (1999): *Standards for educational and psychological Testing*. Washington. APA.
- Arraiz, A. y Sabirón, F. (2005): El portafolios, una herramienta intersubjetiva para la evaluación etnográfica de competencias. *En Actas AIOSP/IAEVG- Internacional Conference*, Lisboa, 14-16 de septiembre.
- Barberá, E. (2005). La evaluación de competencias complejas. *EDUCERE*. Vol. 31.
- Barragán, R. (2005): El portafolio, metodología de evaluación y aprendizaje de cara al nuevo Espacio Europeo de Educación Superior. Una experiencia práctica en la universidad de Sevilla, *Revista Latinoamérica de Tecnología Educativa*, 4(1).
- Bloom, B. et al. (1990): *Taxonomía de los objetivos de la educación: Clasificación de las metas educacionales*. Buenos Aires: El Ateneo.
- Casanova, M. A. (1998): *La Evaluación Educativa*. México, SEP - Cooperación Española.
- Casanova, M. A. (2007): *Manual de Evaluación Educativa*. 9ª ed. Madrid, España, Editorial la Muralla, S. A.
- Centro Nacional de Evaluación para la Educación Superior, (1999): *Manual Técnico*. México, CENEVAL.
- Committee on Test Standards: American Educational Research Association, National Education Association and National Council on Measurements Used in Education (1955): *Technical recommendations for achievement tests*. Washington: The National Education Association.
- Danielson, C. y Leslie, A. (1999): *Una Introducción al uso de portafolios en el aula*. México. Fondo de Cultura Económico.
- Furst, E. J. (1994): Bloom's taxonomy philosophical and educational issues. En L. W. Anderson y L. A. Sosniak (Eds.) *Bloom's taxonomy a frothy-year retrospective*. 93rd Yearbook of the National Society for the Study of Education.
- Gagné, R. M. (1970): *Las condiciones del aprendizaje*. Madrid: Aguilar.
- Gronlund, N. E. (1982): *Elaboración de tests de aprovechamiento*. México: Trillas.
- Haladyna, T. M. (1997): *Writing test items to evaluate higher order thinking*. Boston: Allyn & Bacon.

- Johns, J. (1992): "How professionals view portfolio assessment" en *Reading Research and Instruction*, núm. 32, 1.
- Jornet, J. M. (En prensa): "La evaluación de los aprendizajes universitarios". *Publicaciones de la Universidad de Cádiz*.
- Jornet, J. M.; Sánchez, P. y Leyva, Y. (2008) Dimensiones de clasificación de los procesos de Evaluación Educativa. En Jesús M. Jornet y Yolanda E. Leyva (Coords.) *Conceptos, metodología y profesionalización en la evaluación educativa*. INITE, México.
- Klenowsky, V. (2005): *Desarrollo del portafolios para el aprendizaje y la evaluación: procesos y principios*. Nancea, Madrid.
- Leyva, Y. E., (2009): *Fundamentos de la educación basada en competencias*. (En prensa) México, INITE.
- Leyva, Y. E., y Jornet, J.M. (2006). El perfil del evaluador educativo. *Boletín CENEVAL*, nº 7 (Nov.), págs. 2-6. México: CENEVAL.
- Marzano, R. J. (2000): *Designing a New Taxonomy of Educational Objectives*. Corwin Press, Inc. A Sage Publications Company. California.
- Marzano, R. J. (2006): *Designing a New Taxonomy of Educational Objectives*; 2nd. Ed., Corwin Press, Inc. A Sage Publications Company, California.
- Merrill, M. D. (1994): *Instructional design theory*. Englewood Cliffs, NJ: Educational.
- Miras, M. y Solé, I. (1991): "La evaluación del aprendizaje y la evaluación en el proceso enseñanza-aprendizaje". En Coll, C., et. al. (Comp.) *Desarrollo Psicológico y Educación II. Psicología de la Educación*. España: Alianza Editorial.
- Murphy, R. (2006): Evaluating new priorities for assessment in higher education. En C. Bryan y K. Clegg (Eds.) *Innovative Assessment in Higher Education*. New York. Routledge.
- National Education Association (1993): *Student Portfolios*, Washintong, D. C. National Education Association.
- Polín, L. (1991): "Portfolio Assessment". In *The Writing Notebook*, Núm 8, 3
- Ramos, G. Perales, M. J. y Pérez, M. A. (2009): El concepto de Evaluación Educativa. En Jesús M. Jornet y Yolanda E. Leyva (Coomps.) *Conceptos, metodología y profesionalización en la evaluación educativa*. México, INITE.
- Scriven, M. (1967): The methodology of evaluation: Perspectives of curriculum evaluation. *AERA monograph series on curriculum evaluation*, 1, 39-83. Chicago: Rand-McNally

- Seda-Santana, I. (2002): Evaluación por portafolios. *Revista Latinoamericana de Estudios Educativos*. 1-34.
- Stevens, D. D. & Levi, A. J. (2005): *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning*. Canada, Stylus Publishing, LLC.
- Tenbrink, Terry (1997): *Evaluación: Guía Práctica para Profesores*. Madrid, Narcea S. A. de Ediciones. 255-273.
- Tuckman, B. W. (1979): *Evaluating instructional programs*. Boston: Allyn & Bacon.
- Tristán López, Agustín (1995): "Nuevas capacidades de Kalt para el dictamen de cuestionarios objetivos". *Memoria del I Foro Nacional de Evaluación Educativa*. México: CENEVAL.
- Tristán, Agustín, (1996): "Formulación directa para calcular la Confiabilidad con Alfa de Cronbach", Cuadernos del CENEVAL 96-1.
- Tristán, Agustín (1999) Contribución al estudio del error de medida (parte 1). Notas sobre Evaluación Criterial. CENEVAL / ICI.
- Tristán, Agustín (1999) Contribución al estudio del error de medida (parte 1). Notas sobre Evaluación Criterial. CENEVAL / ICI.
- Yániz, C. y Villardón, L. (2006): *Planificar desde competencias para promover el aprendizaje*. Bilbao, Mensajero.
- Wiggins, G. (1989): A True Test: Toward More Authentic and Equitable Assessment. *Phi Delta Kappan*, 70/9, 703-713.