

# Guía procedimental y metodológica para la elaboración de instrumentos de evaluación

2018

# Índice

<b>I. ¿CON QUÉ EVALUAR?</b>	<b>3</b>
1 SITUACIÓN DE EVALUACIÓN:	3
2 PROCEDIMIENTO DE EVALUACIÓN:	3
3 INSTRUMENTO DE EVALUACIÓN:	3
<b>II. CLASIFICACIÓN DE INSTRUMENTOS DE EVALUACIÓN:</b>	<b>3</b>
2 PRUEBAS ESCRITAS OBJETIVAS:	3
2.1 SELECCIÓN MÚLTIPLE:	5
2.1.1 SUGERENCIAS PARA SU ELABORACIÓN:	5
2.2 PRUEBAS ESCRITAS DE DESARROLLO:	6
2.2.1 PREGUNTAS DE DESARROLLO BREVE:	6
2.2.2 SUGERENCIAS PARA SU ELABORACIÓN:	6
2.2.3 SUGERENCIAS PARA LA REDACCIÓN DE PREGUNTAS DE DESARROLLO BREVE:	6
2.2.4 PREGUNTAS DE DESARROLLO EXTENSO:	7
2.2.5 SUGERENCIAS PARA LA REDACCIÓN DE PREGUNTAS DE DESARROLLO EXTENSO Y ENSAYO.	7
2.2.6 TAXONOMÍA DE MARZANO Y VERBOS SUGERIDOS PARA LA CONSTRUCCIÓN DE ÍTEMS DE SELECCIÓN MÚLTIPLE, DESARROLLO BREVE Y EXTENSO. (UNAM, 2014).	8
3 OBSERVACIÓN DIRECTA:	9
3.1 ESCALA DE APRECIACIÓN:	9
3.1.1 PASOS SUGERIDOS PARA SU CONSTRUCCIÓN:	9
3.2 PAUTA DE COTEJO:	10
3.2.1 PARA QUE LA INFORMACIÓN QUE APORTAN SEA ÚTIL, ES NECESARIO:	10
3.2.2 PASOS SUGERIDOS PARA SU CONSTRUCCIÓN:	10
3.3 RÚBRICA:	10
3.3.1 SU PROCESO DE CONSTRUCCIÓN DEBE CONSIDERAR LOS SIGUIENTES AL MENOS PASOS:	10
3.3.2 RECOMENDACIONES PARA SU APLICACIÓN:	11
3.3.3 ¿CÓMO AUTOEVALUAR LA CALIDAD DE SU RÚBRICA?	11
3.3.4 PAUTAS DE VALIDACIÓN DISCIPLINAR Y TÉCNICA DE UNA RÚBRICA.	12
<b>III. CALIDAD DE LAS EVALUACIONES</b>	<b>13</b>
1 VALIDEZ:	13
1.1 PREGUNTAS SUGERIDAS RESPECTO A LA VALIDEZ DE UNA EVALUACIÓN.	13
2 CONFIABILIDAD DE LA EVALUACIÓN:	13
3 FACTORES QUE AFECTAN LA VALIDEZ Y LA CONFIABILIDAD DE UN PROCESO EVALUATIVO	14
<b>IV. BIBLIOGRAFÍA</b>	<b>15</b>

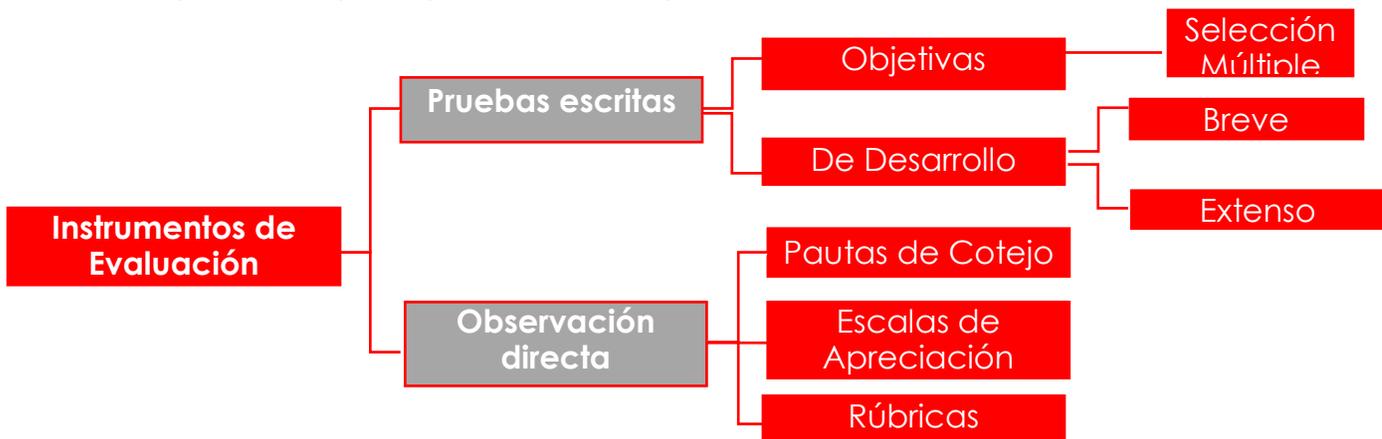
## I. ¿Con qué evaluar?

### Definiciones que considerar:

- 1 **Situación de evaluación:** Es el contexto en el cual se presenta la evaluación, por lo tanto, es planificada e intencionada, ya que es la manifestación de la triple concordancia en una actividad que busca obtener evidencias observables, medibles, calificables y cuantificables del nivel de logro de los resultados de aprendizaje de los estudiantes.
- 2 **Procedimiento de evaluación:** Estrategia que se utiliza para evaluar la actividad que realiza el estudiante y por la cual será evaluado: Aprendizaje Basado en Problemas, Método de Casos, Aprendizaje Basado en Proyectos, Juegos y Simulación, Juego de Roles, Evaluación Clínica Objetiva estructurada (ECO), Aprendizaje + Servicio (A+S), Lectura Previa "clase a clase", Exposición, Salidas a Terrero, Panel de Discusión - Debate, Demostraciones, Disertación, Investigación, Etc.
- 3 **Instrumento de evaluación:** Herramienta que contempla los criterios de evaluación y los indicadores de la tarea que debe realizar el estudiante que permite emitir un juicio sobre su desempeño.

## II. Clasificación de instrumentos de evaluación:

Tradicionalmente, la elaboración de instrumentos de evaluación se basa en dos áreas, por un lado, aquellas clasificadas como pruebas escritas y, por otro lado, aquellas clasificadas como observación directa. Siendo el uso, propósito, información que reporta, estructura interna, entre otras, razones de su distinta clasificación. Metodológicamente utilizaremos el presente esquema para orientar la explicación de cada instrumento.



### 2 Pruebas escritas objetivas:

Test de lápiz y papel altamente estructurado donde la respuesta del estudiante queda manifiesta y graficada por una marca que indica preferencia o selección entre respuestas (alternativas) posibles. Este test se caracteriza por poseer un gran número de ítems – generalmente, los buenos test objetivos tienen entre 35 y 100 ítems-, y por ser completamente objetivos en su corrección (de ahí su nombre) (Pizarro, 2004).

a) Principales ventajas:

- Posee mayor cantidad de ítems, por lo mismo es **más representativo**, al tener un mayor muestreo.
- Cuenta con **mayor validez** de contenido, de construcción y predictiva. Esta segunda ventaja se deriva directamente de la primera; además, por la delicada planificación tanto de la **tabla de especificaciones**, como por la estructuración y forma de los ítems.
- Al ser más válido, es por lo mismo **más confiable**. Esta es una prueba que puede brindar un alto grado de consistencia interna y externa. Da más confianza, y es más digna de crédito –al estar bien elaborada, por supuesto-. El docente no tiene tantas dudas acerca de los datos como los tiene con la prueba oral y de ensayo.
- Es **totalmente objetivo** de corrección. La objetividad de corrección se asegura al corregir los ítems con una plantilla única de corrección. Tal es la razón por la cual esta prueba recibe el nombre de objetiva.
- Implica un **tiempo mínimo por corrección**.
- En resumen, es una **prueba bastante práctica** y cómoda para implementar.

b) Principales desventajas:

- Requiere **bastante tiempo de elaboración**. Sin duda, es la prueba que exige el mayor gasto de tiempo en su elaboración confección de ítems, ordenamiento y estructuración de los ítems, calibración de las respuestas correctas –debe existir una ponderación o idéntica o similar entre las alternativas correctas-, ordenación por contenidos y/o dificultad estimada o comprobada de los ítems, etc.
- Esta prueba no está pensada para medir algunos niveles cognitivos complejos; síntesis, autorregulación o habilidades que de suyo carecen de respuesta única como la expresión escrita, etc.
- En esta prueba, el estudiante no tiene libertad ni de interpretación de los ítems, ni de respuesta. Debe limitarse a seleccionar la alternativa –entre varias- que estime correcta para cada ítem.
- El mayor problema de esta prueba lo constituye, sin duda, el factor azar. El test objetivo es el único test que por forma presenta un azar posible de respuesta. Este oscila entre un 20% y un 50%.

**2.1 Selección múltiple:** Son preguntas objetivas altamente estructuradas que se caracterizan por presentar un enunciado (tronco o base) que brinda la información, y, posibles respuestas a tal enunciado (alternativas); de las cuales es posible elegir la que se estima correcta. Pueden presentar tres, cuatro y cinco alternativas. Los más usados son los ítems de cuatro o de cinco alternativas, pues al presentar un azar menor de respuestas, hacen más difícil los tests que los contienen; por lo mismo aumenta la desviación y con ello la discriminación y la confiabilidad.

### 2.1.1 Sugerencias para su elaboración:

- Evitar que la respuesta correcta sea más larga. Generalmente uno tiende a sobre explicar lo correcto y viceversa. En lo posible haga todas las alternativas de un largo similar.
- **Establecer** la plausibilidad entre el enunciado y todas las alternativas, es decir, que sea una relación válida, aceptable y creíble. La plausibilidad asegura que todas las alternativas aparezcan como correctas para la persona que no sabe; o, que sólo sabe o domina ligeramente los objetivos educativos contrastables.
- Cuando utilice números en los contenidos de las alternativas, proceda a ordenarlas de menor a mayor o viceversa.
- **Evitar** la asociación verbal entre la base y las alternativas, específicamente entre la base y la respuesta correcta. El estudiante, con o sin entrenamiento en tests objetivos, identifica inmediatamente la asociación verbal, o asociación “por resonancia” entre la base y la respuesta correcta.
- **Incluir** en el enunciado debe lo necesario para comprender el sentido de la pregunta y de la respuesta. Razones: pérdida de tiempo, tanto en lectura, como de respuestas; adjuntar datos e informaciones o explicativas, o, que no vienen al caso; confunden al estudiante
- Las palabras que se repiten en todas las alternativas deben formar parte del enunciado.
- Debe existir congruencia gramatical entre todas las alternativas y el enunciado. Las fallas más comunes en la elaboración de ítems objetivos – de cualquier tipo o forma – se deben principalmente a la poca congruencia o concordancia gramatical que presentan los ítems. La congruencia gramatical se debe respetar en cuanto al número (singular o plural), genero (masculino, femenino, neutro), tiempo verbal (presente, pasado, etc.), tipo de término usado (las alternativas o contienen sustantivos, adjetivos, o pueden empezar con un artículo o verbo).
- **Evitar** las opciones “ninguna de las anteriores”, o “todas las anteriores”. Generalmente este tipo de alternativas es usado en tests de matemática y/o de gramática, especialmente ortografía. Sin embargo, no son muy aconsejables debido a que representan casi siempre incongruencia gramatical con el enunciado; generalmente, se pregunta afirmativamente, negando de plano toda posibilidad de elección de esta alternativa; dificultad excesivamente los ítems y hacen poco nítido el mensaje para el alumno; a veces, se los coloca para justificar otra alternativa que ayuda a reducir al azar, o, por falta de imaginación. “todos los anteriores” presenta los mismos problemas; sin embargo, es más usable en el tipo de ítem denominado SM Complejo o con Requisito.

- Asegúrese de que **cada ítem sea independiente**. La respuesta de un ítem no debe condicionar la respuesta de otro ítem del mismo test.

**2.2 Pruebas escritas de desarrollo:** El principio sustantivo de este tipo de preguntas es pesquisar, estimar, medir y evaluar la producción o creación escritas de las respuestas que los estudiantes dan a las pruebas que responden. Tradicionalmente estas pruebas pueden estar constituidas por preguntas de **Desarrollo Breve** y **Desarrollo Extenso**.

**2.2.1 Preguntas de Desarrollo Breve:** Por su naturaleza, estas preguntas junto con poder medir, estimar y evaluar los niveles taxonómicos de los ítems de Selección Múltiple, además pueden abordar habilidades complejas tales como producción, creación, síntesis, elaboración libre y original. Sin embargo, la gran diferencia es que estas preguntas limitan la extensión de las respuestas de los estudiantes, de ahí su nombre.

a) **Principales ventajas:**

- Su aporte más destacable es acompañar y, por tanto, complementar las preguntas de cerradas de Selección Múltiple, pues permiten aportar producción creativa personal

b) **Principales desventajas:**

- Representatividad y validez, pues usualmente estas pruebas presentan pocas preguntas de este tipo.
- La redacción y/o representación de sus preguntas, textos, figuras, gráficos lecturas, etc., debe ser precisa y de alta claridad para permitir la redacción de respuestas muy breves.
- Requieren de mayor tiempo y consenso para leer, corregir y evaluar respuestas.

**2.2.2 Sugerencias para su elaboración:**

- Redactar en forma clara y precisa, cautelando incluir todo lo necesario para que el estudiante comprenda lo que se pregunta y pueda responder apropiada y brevemente.
- Adecuar el nivel de dificultad y del lenguaje al grupo de estudiantes evaluados y al criterio de evaluación respectivo.
- Evitar las redacciones extensas pues dificultan la comprensión del estudiante, y, genera respuestas extensas que no corresponden a este tipo de ítems; sino, al de ensayo.
- No brindar pistas gramaticales (género, número, verbo, estilo, etc.) que ayuden a que el estudiante pueda por ello –y no por su aprendizaje– llegar fácilmente a la respuesta correcta.

**2.2.3 Sugerencias para la redacción de preguntas de desarrollo breve:**

- Quién, qué, cuándo, cuál, a quién, de dónde;
- Defina, cite, nombre, identifique;
- Enumere, haga una lista, ordene;
- Haga un esquema, organice, clasifique;
- Describa, mencione, narre, escriba acerca de; (e) contraste o compare, distinga entre, diferencie;
- Explique, demuestre cómo, demuestre por qué, dé razones que demuestren;
- Discuta, defienda, refute, pruebe, justifique;

- Desarrolle, analice, relacione;
- Sintetice, integre, diseñe, cree, construya, idee;
- Evalúe, juzgue, interprete, emita una opinión fundada.

**2.2.4 Preguntas de Desarrollo Extenso:** Corresponde a aquellos reactivos respondidos por el estudiante mediante una composición de libre elaboración, estilo y originalidad. Permite evaluar niveles taxonómicos y habilidades complejas, tales como capacidad de establecer relaciones, síntesis y análisis entre otros; además pueden ser apreciadas habilidades gramaticales y expresión escrita en general.

**a) Principales ventajas:**

- Favorece la originalidad, pues el estudiante responde lo que quiere y como quiere. Da la libertad para organizar y crear según el criterio del que responde.
- Permite apreciar la capacidad para emitir juicios críticos y de valor.
- Es una prueba que da más seguridad – comparada con la prueba oral – para emitir las respuestas.
- Da las mismas preguntas a todos los estudiantes, y un tiempo posible de respuesta idéntico para el grupo-curso; todo lo cual habilita comparaciones basadas en criterios idénticos.

**b) Principales desventajas:**

- Es la prueba que más gasto de tiempo necesita para corregir y calificar.
- Son difíciles de calificar y por lo general muy subjetivas. Lo cual hace obligatorio acompañarlas de adecuadas instrucciones y una Rúbrica que explicité qué y cómo se evaluará cada respuesta.

**2.2.5 Sugerencias para la redacción de preguntas de Desarrollo extenso y ensayo.**

- **Número de Ítems:** Las pruebas de ensayo generalmente tienen una cantidad de ítems que oscila entre 3 y 12, al establecer un número mayor de ítems es más probable que se aumente la validez (al aumentar la representatividad), y como consecuencia de lo anterior, puede aumentar la confiabilidad.
- **Redacción de los Ítems:** Existe una regla empírica en la formulación de las preguntas de las pruebas de ensayo, que consiste en que, si la pregunta es larga, la respuesta será corta; ya que al redactar preguntas largas se especifica mejor la tarea que se requiere que los estudiantes cumplan. Por lo mismo, si hay un aumento de precisión en las preguntas, debe existir igual aumento en las respuestas; las que a su vez son más cortas.
- No utilizar demasiado las siguientes palabras: Qué, Quién, Cuándo, Dónde, Mencione, Enumere; ya que estas palabras exigen reproducción de información (memorización o recuerdo). Debe usarlas cuando sus objetivos correspondan a ese nivel.
- Se sugiere que las preguntas de la prueba de ensayo tengan una dificultad intermedia y que estén de acuerdo con el grupo y sus respectivos Resultados de Aprendizaje y criterios de evaluación.
- **Adecuar el tiempo posible de respuesta.** Generalmente los estudiantes demoran entre 2 y 4 veces más de lo que el docente se demora en contestar la misma prueba. Es recomendable probar empíricamente esta estimación del tiempo. Contesté usted primero la prueba, luego su ayudante u otro profesor de la misma especialidad. Los tiempos promedio de respuesta multiplíquelo por 4, de tal suerte que ese sería el tiempo máximo posible de respuesta.

- Especificar el valor o puntuación de cada pregunta. Sugerimos usar la **graduación 1 – 3**, así las preguntas con mayor **abstracción y dificultad taxonómica** tendrán **ponderación tres**, y las de menor dificultad o ponderación, tendrán **ponderación uno**, existiendo, consecuentemente a lo anterior, el punto intermedio en la **ponderación dos**.
- Elaborar una rúbrica de corrección escrita que contenga puntos o aspectos estimados correctos, mediante correctos y deficientes. La pauta puede tener solamente puntos específicos estimados correctos. No puede ni debe estar redactada la respuesta correcta por pregunta; ya que en esta prueba no hay una sola respuesta correcta; o, mejor dicho, no existe una sola forma de respuesta correcta.

### 2.2.6 Taxonomía de Marzano y verbos sugeridos para la construcción de ítems de Selección múltiple, desarrollo breve y extenso. (UNAM, 2014).

Nivel	Clasificación	Definición	Verbos sugeridos
1	Conocimiento/ recuerdo/ (recuperación)	Usa una variedad efectiva de estrategias de razonamiento complejas. Traduce ediciones efectivamente y situaciones a tareas manejables con propósitos claros.	<ul style="list-style-type: none"> <li>▪ Reconocer</li> <li>▪ Recordar</li> <li>▪ Ejecutar</li> <li>▪ Nombrar</li> <li>▪ Identificar</li> <li>▪ Realizar un procedimiento</li> </ul>
2	Comprensión	Usa una variedad de técnicas reunidas de información y recursos efectivos. Interpreta y sintetiza la información efectivamente. Asesora exactamente el valor de la información. Reconoce dónde y cómo los proyectos pueden beneficiarse con información adicional.	<ul style="list-style-type: none"> <li>▪ Integrar</li> <li>▪ Simbolizar</li> <li>▪ Ubicar</li> <li>▪ Sintetizar</li> <li>▪ Representar</li> <li>▪ Presentar</li> </ul>
3	Análisis	Expresa ideas claras. Comunica efectivamente con diversas audiencias. Comunica efectivamente en una variedad de formas. Comunica efectivamente en una variedad de propuestas. Crea productos con calidad.	<ul style="list-style-type: none"> <li>▪ Relacionar</li> <li>▪ Asociar</li> <li>▪ Identificar</li> <li>▪ Analizar</li> <li>▪ Generalizar</li> <li>▪ Clasificar</li> <li>▪ Analizar errores</li> <li>▪ Especificar</li> </ul>
4	Utilización del conocimiento	Trabaja hacia logros de las metas del grupo. Usa habilidades interpersonales efectivamente. Contribuye al mantenimiento de un grupo. Realiza una variedad de roles efectivamente.	<ul style="list-style-type: none"> <li>▪ Tomar decisiones</li> <li>▪ Resolver problemas</li> <li>▪ Experimentar</li> <li>▪ Investigar</li> <li>▪ Utilizar</li> <li>▪ Generar</li> <li>▪ Evaluar</li> <li>▪ Conducir</li> </ul>
5	Meta cognición	Autorregulación Es consciente del pensamiento propio. Hace planes efectivos. Es consciente y usa los recursos necesarios. Es sensible a la retroalimentación. Evalúa la efectividad de sus propios actos.	<ul style="list-style-type: none"> <li>▪ Trazar</li> <li>▪ Lograr</li> <li>▪ Conseguir (algo que se pretende)</li> <li>▪ Monitorear</li> <li>▪ Evaluar</li> <li>▪ Especificar</li> <li>▪ Determinar (qué tan bien se ha hecho algo y qué tan efectivo es)</li> </ul>
6	Sistema interno (Consciencia del ser)	Compromete intensivamente en tareas aun cuando las respuestas o soluciones no aparecen inmediatamente. Empuja al límite del propio conocimiento y habilidades. Genera, confía y mantiene sus propios estándares de evaluación. Genera nuevos puntos de vista a situaciones fuera de límite de los estándares convencionales.	<ul style="list-style-type: none"> <li>▪ Mejorar</li> <li>▪ Calificar (el propio desempeño)</li> <li>▪ Evaluar (la importancia del conocimiento adquirido; la eficacia, las emociones y la motivación)</li> </ul>

### 3 Observación directa:

Los instrumentos de evaluación por observación son aquellos que permiten evaluar la ejecución de una determinada tarea solicitada al estudiante, mediante la observación directa de la evidencia de desempeño que éste realiza.

Este tipo de instrumentos son de gran utilidad, pues permiten evaluar en tiempo real la ejecución del desempeño que se ha solicitado al estudiante a partir de una **situación de evaluación** planificada, por lo tanto, promueve una entrega de resultados inmediata y, por consiguiente, favorece la retroalimentación oportuna y objetiva respecto a la calidad de la evidencia o producto ejecutado, expuesto o entregado.

La clasificación utilizada usualmente de los instrumentos de evaluación por observación, en el contexto de los enfoques por competencias, es: **escalas de apreciación, rúbricas y pautas de cotejo**, siendo la finalidad y momento de uso, criterios para diferenciar a cada uno.

**3.1 Escala de Apreciación:** El proceso de evaluación por juicio se llama apreciación y juicio se define como la formación de una opinión o de una decisión. Ahora bien, cuando una opinión afecta a la evaluación, ésta se torna más subjetiva.

Así, como su nombre lo indica, estas **escalas permiten apreciar**, un **desempeño** según la frecuencia o calidad en que se manifiestan aquellos **descriptores** que lo componen, entendidos éstos a su vez como las partes, pasos o conductas que lo componen. Su utilidad está en permitir entregar una estimación del proceso de adquisición de una determinada competencia, por lo mismo su uso es recomendado en evaluaciones diagnósticas o formativas.

Estas escalas consisten en una lista de características o rasgos acompañados de una o varias escalas con las cuales se podrá establecer el grado o medida en que dicha característica o rasgo se presenta en un sujeto.

#### 3.1.1 Pasos sugeridos para su construcción:

- Definir el desempeño a observar.
- Determinar las características del desempeño y los aspectos que permiten observar su realización, logro o calidad.
- Seleccionar los aspectos que se van a observar.
- Explicitar las instrucciones para utilizarla
- Enumerar los rasgos importantes de la conducta o comportamiento a evaluar.
- Indicar el tipo de escala a utilizar (Siempre- Frecuentemente-Ocasionalmente- Casi Nunca – Nunca).
- Asignar puntaje (5=Siempre; 4=Frecuentemente; 3=Ocasionalmente; 2=Casi Nunca; 1=Nunca).

**3.2 Pauta de Cotejo:** Corresponde a una lista de palabras, frase u oraciones que señalan con mucha especificidad, ciertas tareas, acciones, procesos, productos de aprendizaje, conductas positivas o negativas. Frente a cada palabra, frase u oración, denominados genéricamente como descriptores, se desagrega un determinado desempeño, posee una escala de valoración dicotómica que evalúa la ausencia o presencia de cada descriptor, por tanto, es utilizado preferentemente al final del proceso de enseñanza aprendizaje, pues permite certificar la adquisición de una determinada competencia.

### 3.2.1 Para que la información que aportan sea útil, es necesario:

- Describir precisamente las acciones específicas seleccionadas.
- Especificar el orden o secuencia en que se realice una serie de acciones o procedimientos en caso de que ello sea importante.
- El conjunto de los aspectos registrado como presentes o ausentes, sólo aportan una información descriptiva y muy pocas veces corresponderá a una forma de medición en el sentido estricto de la palabra.

### 3.2.2 Pasos sugeridos para su construcción:

- Definir el desempeño a observar.
- Determinar las características deseables del desempeño
- Seleccionar los aspectos que se van a incluir.
- Explicitar las instrucciones para utilizarla.
- Colocar los pasos específicos en orden consecutivo
- Indicar tipo de escala a utilizar (sí –no; logrado-no logrado; presencia – ausencia; etc.).
- Asignar puntajes.

**3.3 Rúbrica:** considerada el instrumento de observación por excelencia al momento de evaluar desempeños, permite objetivar aquello que se espera sea desarrollado por los estudiantes. Está compuesto por dimensiones o criterios señalados como aspectos considerados importantes de evaluar, niveles de logro con una escala de valoración decreciente y descriptores de cada dimensión según la escala de valoración seleccionada. De esta forma es posible obtener información tanto para profesores y estudiantes respecto al grado de desarrollo o calidad de la evidencia de desempeño realizada.

La rúbrica puede ser utilizada para objetivar el qué evaluar, retroalimentar adecuadamente el aprendizaje de los estudiantes, promover la auto y co-evaluación, propender a la coherencia en la calificación, democratizar el proceso de evaluación.

### 3.3.1 Su proceso de construcción debe considerar los siguientes al menos pasos:

- Definir los criterios de evaluación a medir.
- Plantear una situación de evaluación.
- Reflexionar sobre el óptimo resultado o evidencia de desempeño.
- Proponer un título que sintetice lo que se evaluará.
- Vincular el desempeño posible al desarrollo cognitivo de los estudiantes.
- Expresar claramente el desempeño evaluar.
- Elaborar una lista de los aspectos, etapas o rasgos característicos del desempeño a evaluar (criterios o dimensiones). Se deben seleccionar sólo aquellos aspectos que sean independientes entre sí, es decir, que no estén relacionados de manera que el puntaje o ponderación de uno influyan en el otro.
- Recomendable es que no sean más de cuatro o cinco criterios.

- Organizar los criterios de acuerdo con el tipo de desempeño que se quiere evaluar. Secuencia lógica o procedimental.
- Decidir el número de categorías que se empleará.
- Definir la escala de valoración de desempeño, la cual deberá ser en orden decreciente.
- Describir cada categoría en forma exhaustiva para cada criterio, utilizando un lenguaje preciso y sin ambigüedades.
- Cotejar o validar el instrumento con un colega de área, con el fin de recibir comentarios o retroalimentación y así verificar si efectivamente evalúa lo que declara evaluar.

### **3.3.2 Recomendaciones para su aplicación:**

- Aplicar en forma piloto, para determinar eventuales ajustes.
- Socializar previamente con el grupo al cual será aplicada, de esta forma se informa oportunamente qué se espera del desempeño en relación a la situación de evaluación propuesta.
- Explicar el sentido y finalidad del instrumento, recoger y aclarar dudas.
- Entregar resultados a los estudiantes en forma oportuna, a fin de retroalimentar el desempeño de cada uno en relación a las dimensiones evaluados y los niveles de logro establecido.

### **3.3.3 ¿Cómo autoevaluar la calidad de su rúbrica?**

- Evaluar que los criterios sean suficientes y necesarios para observar/medir el desempeño de todos los estudiantes con respecto a los criterios de evaluación seleccionados.
- Cautelar que las expectativas sobre el desempeño de los estudiantes sean claras y que los niveles de desempeño se encuentren redactados en un lenguaje comprensible para los alumnos, padres y apoderados.
- Evitar la formulación de los niveles de desempeño se evitan expresiones ambiguas o difíciles de definir tales como "actitud positiva", "estilo elegante", "calidad emergente", etc.
- Evitar que en los niveles inferiores el uso de términos negativos. Por ejemplo "no domina la ortografía", "no tiene ninguna habilidad para la expresión oral", "no sabe utilizar el color", etc.

Complementariamente a lo anterior, se sugiere la utilización de la siguiente pauta de validación de su rúbrica.

### 3.3.4 Pautas de validación disciplinar y técnica de una Rúbrica.

**Objetivo:** Las siguientes pautas permiten validar la rúbrica construida correspondiente cada electivo, según su juicio experto, desde el ámbito disciplinar y técnico. Para lo cual agradecemos su colaboración.

<b>Asignatura:</b>	
<b>Sede Elaboradora</b>	
<b>Sede validadora</b>	
<b>Validador/a</b>	
<b>Fecha</b>	

#### I. Validación disciplinar.

Los aspectos que solicitamos valide, para cada dimensión de la Rúbrica, como evaluador experto son:

1. **Pertinencia:** Concordancia entre cada dimensión, su definición y la ponderación asignada.
2. **Validez de contenido:** Ajuste a criterio, es decir, los niveles de logro (NL) permiten evaluar el desempeño de cada dimensión.

Según lo anterior, revise cada dimensión y sus indicadores de logro indicando su acuerdo o desacuerdo, justificando su decisión.

DIMENSIÓN	Pertinencia		Validez de Contenido de los niveles de logro								Justificación
			NL 4		NL 3		NL 2		NL 1		
	Sí	No	Sí	No	Sí	No	Sí	No	Sí	No	
1.											
2.											
3.											
4.											
5.											
6.											

Mediante la pauta de cotejo<sup>1</sup>, evalúe la calidad técnica de las dimensiones y descriptores de los niveles de desempeño.

°	Características deseables	Sí	No	Justificación
1	Las dimensiones son suficientes y necesarios para juzgar el desempeño de todos los estudiantes con respecto los Resultados de Aprendizaje seleccionados.			
2	Los niveles considerados representan el rango completo de conocimientos, habilidades y actitudes que constituyen los Resultados de Aprendizaje del proceso de enseñanza aprendizaje.			
3	Las expectativas sobre el desempeño de los estudiantes son claras. Los niveles de desempeño se encuentran redactados en un lenguaje comprensible para todos los estudiantes,			
4	Se ha considerado al menos un nivel sobre el desempeño mínimo esperado, para ejemplificar los desempeños superiores.			
5	En la definición de los niveles de desempeño se evita la omisión o existencia de niveles implícitos, que dejan a la audiencia el grado o tipo de desempeño que existe entre los niveles explícitos.			
6	Las dimensiones y niveles de desempeño se encuentran formulados para permitir la retroalimentación continua y no sólo una calificación de desempeño aislada.			
7	Los rasgos definidos son claros considerando en forma específica desde el nivel que señala el "menor" desarrollo hasta el que involucra el "mayor" desarrollo.			
8	Los niveles de desempeño se encuentran en todas las dimensiones.			
9	En los niveles inferiores no se utilizan sólo términos negativos. Por ejemplo "no domina la ortografía", "no tiene ninguna habilidad para la expresión oral", "no sabe utilizar el color", etc.			
10	Se utilizan categorías absolutas y no relativas al nivel máximo de desempeño encontrado en el grupo.			
11	En la formulación de los niveles de desempeño se evitan expresiones ambiguas o difíciles de definir tales como "actitud positiva", "estilo elegante", "calidad emergente", etc.			
12	En la formulación de los niveles de desempeño se distingue claramente entre los niveles intermedios y los niveles que aún requieren reforzamiento, evitando expresiones como "promedio", "sobre el promedio", "regular", etc.			

<sup>1</sup> Adaptada de HIMMEL, E., OLIVARES M.A. ZABALZA J. (1999) **Hacia una Evaluación Educativa. Aprender para evaluar y evaluar para aprender.** Volumen I. conceptos actuales sobre la evaluación del aprendizaje escolar. Ed. MINEDUC-Pontificia Universidad Católica de Chile, Santiago)

### III. Calidad de las evaluaciones

Se estima muchas veces que la validez y fiabilidad poseen las características de estar o no estar, sin embargo, es necesario aclarar que éstas se presentan en diferentes grados (Prieto & Delgado, 2010).

#### 1 Validez:

La definición más clásica proviene de Anastasi (1998) "validez es lo que la prueba mide y qué tan bien lo hace". Ej. Que un Test de Inteligencia mida inteligencia y no memoria.

##### 1.1 Preguntas sugeridas respecto a la validez de una evaluación.

1. ¿Se evalúa lo que se espera que los estudiantes aprendan o sepan desempeñarse?

Validez de constructo.

Factores que afectan la validez de constructo:

- Diseño de la evaluación.
- Definición de indicadores de logro.
- Redacción de instrucciones.
- Caracterización del desempeño.

2. ¿Corresponde lo que se evalúa a las actividades realizadas durante el proceso de enseñanza y aprendizaje? Validez Instruccional.

- Factores que afectan la validez instruccional.
- Considerar la perspectiva del estudiante.
- Planificación de la evaluación.
- Explicitar los propósitos de la evaluación anticipadamente.

3. ¿Las consecuencias que tiene la evaluación son coherentes con el propósito de la misma? Validez consecuencial.

Factores que afectan la validez consecuencial.

- Definir propósitos.
- Identificación de la evidencia.
- Uso de la evidencia.

#### 2 Confiabilidad de la evaluación:

Medida en que se puede tener confianza en la precisión de las informaciones que sirven de fundamento para emitir un juicio sobre los aprendizajes logrados.

Ejemplificando desde la experiencia, este concepto asume que, si ocupamos un instrumento para estimar el valor de un objeto, balanza para pesar fruta, y si las distintas mediciones realizadas en similares condiciones varían, entonces se considera que aquellas mediciones son inestables, inconsistentes y, por tanto, poco fiables (Prieto & Delgado, 2010).

Factores que afectan la confiabilidad de una evaluación:

- Limitación por la muestra de desempeños seleccionados.
- Insuficiencia de la cantidad.
- Procedimientos con falta de claridad.
- Imprecisión de criterios.

### 3 Factores que afectan la validez y la confiabilidad de un proceso evaluativo

Etapas	Factores
<b>1. Diseño y construcción</b>	<ul style="list-style-type: none"> <li>▪ Poco manejo de la disciplina.</li> <li>▪ Mal planteamiento de objetivos o situaciones de evaluación.</li> <li>▪ Mala ordenación de las situaciones de evaluación o preguntas.</li> <li>▪ Poca claridad de las instrucciones.</li> <li>▪ Reactivos, ítems o preguntas mal planteados</li> <li>▪ Poca coherencia con la metodología empleada para el logro del resultado de aprendizaje planificado.</li> <li>▪ Poca coherencia en la forma de ponderar los reactivos, ítems o preguntas.</li> <li>▪ Poca representatividad de los aspectos evaluados respecto del dominio.</li> <li>▪ Poca coherencia con los criterios a evaluar.</li> <li>▪ Formato de procedimiento de evaluación poco adecuado a los objetivos o al tipo de estudiantes que se quiere evaluar.</li> <li>▪ Falta de claridad respecto de los criterios de evaluación por parte del profesor.</li> </ul>
<b>2. Aplicación</b>	<ul style="list-style-type: none"> <li>▪ Poca claridad de las instrucciones.</li> <li>▪ Nivel de ansiedad de los estudiantes.</li> <li>▪ Poca o ninguna claridad en los estudiantes.</li> <li>▪ Nivel de ansiedad de los estudiantes.</li> <li>▪ Poca o ninguna claridad en los estudiantes, respecto de los criterios con que se va a evaluar.</li> <li>▪ Desconocimiento de parte de los estudiantes del tipo de procedimiento a aplicarse.</li> <li>▪ Dificultades provenientes del contexto: mala infraestructura, falta de material, tiempo, otros.</li> <li>▪ Horas inadecuadas de aplicación.</li> <li>▪ Formatos aplicados distintos a los propuestos.</li> <li>▪ Aplicación por parte de personas desconocidas para los estudiantes.</li> </ul>
<b>3. Corrección</b>	<ul style="list-style-type: none"> <li>▪ No uso de pauta de corrección.</li> <li>▪ Falta de criterios de evaluación claramente establecidos.</li> <li>▪ Corrección en horas o ambientes inadecuados.</li> <li>▪ Cansancio.</li> <li>▪ Efecto halo.</li> <li>▪ Prejuicios del profesor/a.</li> <li>▪ Variados profesores que corrigen</li> </ul>
<b>4. Transformación a calificación.</b>	<ul style="list-style-type: none"> <li>▪ Poca coherencia de la escala a usar con el o los Resultados de Aprendizaje a evaluar.</li> <li>▪ Poca claridad respecto de criterios de aprobación del procedimiento en conjunto.</li> <li>▪ Desconocimiento, ya sea general o particular, de resoluciones, instructivos o reglamentos asociados.</li> </ul>

#### **IV. Bibliografía**

Anijovic, R. La evaluación significativa. Buenos Aires: PAIDÓS. (2010).

Pizarro, R. Manual de Elaboración de Ítemes objetivos de selección múltiple y preguntas abiertas para el segundo estudio regional comparativo y explicativo (SERCE). Santiago: UNESCO. (2004).

Tornimbeni, S. Introducción a la Psicometría. Buenos Aires: PAIDÓS. (2008).

UNAM. Manual para la elaboración de reactivos bajo el enfoque por competencias. Veracruz, México. (2014).