



**Mi Universidad**

**LIBRO**

*EVALUACION EDUCATIVA*

*LICENCIATURA EN CIENCIAS DE LA EDUCACION*

*Quinto Cuatrimestre*

***ENERO- ABRIL***

---

## Marco Estratégico de Referencia

---

### Antecedentes históricos

Nuestra Universidad tiene sus antecedentes de formación en el año de 1979 con el inicio de actividades de la normal de educadoras “Edgar Robledo Santiago”, que en su momento marcó un nuevo rumbo para la educación de Comitán y del estado de Chiapas. Nuestra escuela fue fundada por el Profesor Manuel Albores Salazar con la idea de traer educación a Comitán, ya que esto representaba una forma de apoyar a muchas familias de la región para que siguieran estudiando.

En el año 1984 inicia actividades el CBTiS Moctezuma Ilhuicamina, que fue el primer bachillerato tecnológico particular del estado de Chiapas, manteniendo con esto la visión en grande de traer educación a nuestro municipio, esta institución fue creada para que la gente que trabajaba por la mañana tuviera la opción de estudiar por las tardes.

La Maestra Martha Ruth Alcázar Mellanes es la madre de los tres integrantes de la familia Albores Alcázar que se fueron integrando poco a poco a la escuela formada por su padre, el Profesor Manuel Albores Salazar; Víctor Manuel Albores Alcázar en julio de 1996 como chofer de transporte escolar, Karla Fabiola Albores Alcázar se integró en la docencia en 1998, Martha Patricia Albores Alcázar en el departamento de cobranza en 1999.

En el año 2002, Víctor Manuel Albores Alcázar formó el Grupo Educativo Albores Alcázar S.C. para darle un nuevo rumbo y sentido empresarial al negocio familiar y en el año 2004 funda la Universidad Del Sureste.

La formación de nuestra Universidad se da principalmente porque en Comitán y en toda la región no existía una verdadera oferta Educativa, por lo que se veía urgente la creación de una institución de Educación superior, pero que estuviera a la altura de las exigencias de los jóvenes

que tenían intención de seguir estudiando o de los profesionistas para seguir preparándose a través de estudios de posgrado.

Nuestra Universidad inició sus actividades el 18 de agosto del 2004 en las instalaciones de la 4ª avenida oriente sur no. 24, con la licenciatura en Puericultura, contando con dos grupos de cuarenta alumnos cada uno. En el año 2005 nos trasladamos a nuestras propias instalaciones en la carretera Comitán – Tzitol km. 57 donde actualmente se encuentra el campus Comitán y el corporativo UDS, este último, es el encargado de estandarizar y controlar todos los procesos operativos y educativos de los diferentes campus, así como de crear los diferentes planes estratégicos de expansión de la marca.

## **Misión**

Satisfacer la necesidad de Educación que promueva el espíritu emprendedor, aplicando altos estándares de calidad académica, que propicien el desarrollo de nuestros alumnos, Profesores, colaboradores y la sociedad, a través de la incorporación de tecnologías en el proceso de enseñanza-aprendizaje.

## **Visión**

Ser la mejor oferta académica en cada región de influencia, y a través de nuestra plataforma virtual tener una cobertura global, con un crecimiento sostenible y las ofertas académicas innovadoras con pertinencia para la sociedad.

## Valores

- Disciplina
- Honestidad
- Equidad
- Libertad

## Escudo



El escudo del Grupo Educativo Albores Alcázar S.C. está constituido por tres líneas curvas que nacen de izquierda a derecha formando los escalones al éxito. En la parte superior está situado un cuadro motivo de la abstracción de la forma de un libro abierto.

## Eslogan

“Mi Universidad”

## ALBORES



Es nuestra mascota, un Jaguar. Su piel es negra y se distingue por ser líder, trabaja en equipo y obtiene lo que desea. El ímpetu, extremo valor y fortaleza son los rasgos que distinguen.

---

## Evaluación Educativa

---

### OBJETIVO GENERAL

Los alumnos lograran motivarse por el conocimiento de la evaluación, a través de los temas tratados.

### Criterios y procedimientos de evaluación y acreditación:

Trabajos escritos	30%
Actividades áulicas	20%
Examen	50%
Total	100%
Escala de calificaciones	7-10
Mínima aprobatoria	7

## **UNIDAD I**

### **MEDICION, EVALUACION Y DISEÑO DE PRUEBAS: ELEMENTOS PREVIOS**

- I.1 Historia y desarrollo de la medición y evaluación educativa
- I.2 Medición, evaluación y diseño de pruebas
- I.3 Estructura del libro
- I.4 Análisis de datos en la medición educativa
- I.5 Escalas de medidas
- I.6 El análisis cuantitativo
- I.7 Interpretación normativa de puntuaciones
- I.8 El análisis cuantitativo
- I.9 Teorías sobre las técnicas de obtención de información
- I.10 Teoría clásica de Tests
- I.11 Críticas y alternativas a la teoría clásica de Tests

## **UNIDAD II**

### **ELABORACION DE INSTRUMENTOS DE MEDICION Y EVALUACION EDUCATIVA**

- 2.1 Clasificación de los instrumentos de medición y evaluación educativa
- 2.2 Pruebas de rendimiento de ensayos
- 2.3 Pruebas objetivas de rendimiento de test
- 2.4 Escalas de actitud
- 2.5 Técnicas sociométricas
- 2.6 Cuestionarios
- 2.7 Entrevistas
- 2.8 Instrumentos de observación sistemática
- 2.9 Traducción de los instrumentos de medición y evaluación educativa
- 2.10 La evaluación educativa
- 2.11 Naturaleza de la evaluación



- 2.12 Concepciones evaluativas
- 2.13 Funciones de la evaluación
- 2.14 Tipos de juicios evaluativos
- 2.15 Objetivos, indicadores y estándares
- 2.16 La evaluación como proceso
- 2.17 Componentes del proceso evaluativo
- 2.18 La metaevaluación
- 2.19 Algunos tópicos de interés acerca de la evaluación
- 2.20 Ámbitos básicos de la evaluación educativa

## **UNIDAD III**

### **LA EVALUACION DE LOS APRENDIZAJES DE LOS ALUMNOS**

- 3.1 Cambios fundamentales en la evaluación de los aprendizajes de los alumnos
- 3.2 Modelos de evaluación de los aprendizajes
- 3.3 Procedimientos tradicionales de evaluación de los aprendizajes
- 3.4 Procedimientos alternativos de evaluación de los aprendizajes
- 3.5 A modo de conclusión
- 3.6 Modelos de evaluación alternativa: La evaluación en el contexto del ABP y del portafolios
- 3.7 Modelos alternativos de evaluación de los aprendizajes, principios caracterizadores
- 3.8 La evaluación en el contexto de la metodología del Aprendizaje Basado en Problemas
- 3.9 Elementos críticos del ABP
- 3.10 La evaluación por portafolios
- 3.11 La validez y fiabilidad en los enfoques alternativos de la medición y evaluación educativa
- 3.12 La evaluación de referencia criterial versus la referencia normativa
- 3.13 Origen, desarrollo y naturaleza de las pruebas de referencia criterial
- 3.14 Características técnicas de las pruebas de referencia criterial, elementos diferenciales
- 3.15 La evaluación referida a estándares

3.16 La evaluación basada en ejecuciones

## **UNIDAD IV**

### **LAS COMPETENCIAS Y SU EVALUACION**

4.1 Introducción

4.2 Conceptualizando el termino competencia

4.3 Conocimiento, capacidad, habilidad, competencia

4.4 Tipología de las competencias

4.5 La evaluación de las competencias, elementos previos

4.6 A modo de síntesis

4.7 Epilogo: un nuevo marco conceptual para la medición y evaluación educativa

4.8 Cambio de paradigma en la medición y evaluación educativa

4.9 Aceptaciones en torno al concepto de medición y evaluación educativa. En busca de una definición comprensiva del termino

4.10 En torno a la validez

4.11 En torno a la fiabilidad

4.12 En torno a la naturaleza de los aprendizajes

4.13 En torno a la lógica de la evaluación

4.14 Criterios de calidad alternativos para las pruebas de ejecución

4.15 El factor ético

## UNIDAD I

# MEDICION, EVALUACION Y DISEÑO DE PRUEBAS: ELEMENTOS PREVIOS

### 1.12 Historia y desarrollo de la medición y evaluación educativa

En un primer estadio, la evaluación se entendía como una medición valorada hasta el punto hasta el punto de utilizarse de forma perfectamente intercambiable los vocablos de medición y evaluación educativa.

El comienzo formal de la utilización de la medida en el campo de las ciencias humanas puede situarse en los trabajos de R. Fechner, quien, en la segunda mitad del siglo XIX, introdujo la medición psicofísica, primero en los laboratorios de psicología experimental, y después, asociada al estudio de las diferencias individuales de donde surge la psicometría.

Fue F. Galton, quien introdujo la expresión “test mental”, en un primer intento de clasificar a los individuos a partir de una inteligencia promedio, y por la necesidad de tipificar las condiciones adecuadas para experimentar a los sujetos y poder compararlos entre sí.

Para analizar los datos obtenidos en el laboratorio antropométrico del Museo de South Kensington, Galton recurrió a métodos estadísticos y determino promedios y medidas de varianza. Recurrió a la correlación como medida de asociación para solucionar el problema de descubrir la magnitud de semejanza entre las características individuales de los padres y los hijos. Fue así el primero en aplicar los conceptos estadísticos de curva normal, media, mediana, varianza y correlación de datos psicopedagógicos.

Aunque no profundizó en la metodología estadística, ayudó a K. Pearson a desarrollar los métodos estadísticos asociados al estudio de las diferencias individuales, K. Pearson formalizó

el coeficiente de correlación múltiple y sentó las bases de gran parte de la estadística multivariada.

Los estudios de F. Galton sobre las diferencias individuales se extendieron rápidamente por Alemania, Estados Unidos y Francia, entre otros países.

En Alemania, en el año 1896, F. Keaepelin creó un laboratorio psicológico en el que se midieron los tiempos de reacción de la memoria, la percepción, la discriminación y la fatiga.

En Estados Unidos las pruebas mentales recibieron la influencia de la obra realizada en Europa. La figura más relevante fue J. McKeen Cattell, discípulo de F. Galton, quien realizó pruebas mentales y físicas en el Columbia College.

La contribución de Francia al campo de la medida de las diferencias individuales puede centrarse en A. Binet, quien construyó las primeras escalas para la medida de la inteligencia. A. Binet, en colaboración con H. Simon, elaboró y publicó en 1905 una escala métrica de la inteligencia.

Sin embargo, a principios del siglo XX el principal reto en el campo de la investigación pedagógica era crear tests propiamente escolares y elaborar escalas de comprobación de conocimientos, ya que las pruebas de laboratorio y los tests psicofísicos de hasta entonces resultaron insuficientes para el estudio de los problemas escolares; además, los tests mentales no cubrían las áreas propias del rendimiento escolar.

Aunque se considera que J. M. Rice fue el iniciador de la medición en pedagogía, ésta no se estableció con plena vigencia hasta 1904, con la publicación de "*Introduction to the theory of mental and social measurement*", de R.L. Thomdike.

A partir de 1910, la línea iniciada por J.M. Rice en el campo de la educación se desarrolló rápidamente, con la aparición de escalas de redacción, de ortografía y de cálculo aritmético. Más tarde se crearon las baterías de rendimiento, con la publicación de la primera edición de

Los Stanford Achievement Tests, en 1923, que definieron muchas de las características de la actual medición educativa.

La línea abierta por los tests objetivos de rendimiento gozó de gran aceptación en Estados Unidos, donde se establecieron numerosos programas estatales y nacionales, que en 1947 se fundieron en el Educational Testing Service, el cual elaboró a partir de entonces la mayoría de las pruebas de rendimiento en universidades, escuelas profesionales, organismos gubernamentales y otras instituciones estadounidenses.

Las dos guerras mundiales estimularon considerablemente el movimiento en pro de la creación de tests de todo tipo al servicio de la comunidad. Pero el fervor por el testing decreció a partir de 1950, e incluso surgieron algunos movimientos que se mostraron hipercríticos con dichas prácticas.

### **1.13 Medición, Evaluación y diseño de pruebas**

A continuación vamos a analizar por separado y de forma introductoria y crítica cada uno de estos conceptos que son los que darán estructura y contenido a este texto. En esta primera aproximación buscamos ante todo promover interrogantes que proponer respuestas; será a lo largo del manual que trataremos de estructurar los contenidos de forma que nos ayuden a generar praxis resolutorias a nuestros problemas de medición y evaluación.

#### *La medición*

La pretensión básica de la medición es tratar de cuantificar los atributos que están presentes en objetos o individuos. Con ello se intenta objetivar nuestros juicios valorativos respecto de los mismos y facilitar su estudio y conocimiento.

Con la medición tratamos de hacer comparables los individuos en relación a aquellas propiedades que previamente hemos analizado. La medición empezó a aplicarse a las

Humanidades a partir del siglo XIX cuando se intentó introducir este tipo de conocimientos en la misma metáfora en la que ya estaban situadas las denominadas materias científicas. Medir era la condición previa necesaria para poder aplicar el método científico a cualquier área específica de conocimiento que pretendiera conseguir estatus científico.

La Medición se basa en las variaciones observadas entre los sujetos respecto de alguna característica y pueden tener carácter cualitativo y se suelen expresar mediante las distintas categorías pertenecientes a algún sistema de clasificación o bien en otras ocasiones son susceptibles de cuantificación y podemos hacerlo en términos de frecuencia, cantidad o grado.

En cualquier caso, cuando nos proponemos describir alguna característica de un conjunto de sujetos debemos proceder en primer lugar a especificar o definirla y posteriormente a desarrollar un conjunto de operaciones o métodos de medida que nos permita describir a cada uno de los sujetos en términos de la mencionada propiedad. Cuando nos servimos de algún tipo de descripción cualitativa estamos utilizando, como forma de expresión de la característica, un sistema de clasificación. Cuando lo hacemos mediante una descripción cuantitativa usamos un sistema de medición.

La medición supone el uso de números, que son valores que proveen de descripciones de carácter cuantitativo y pueden ser manipulados a fin de obtener información con valor añadido respecto de los sujetos. Consecuentemente, medir es un proceso por el cual asignamos numerales a las cosas de acuerdo a unas reglas determinadas.

Es con Glaser con quien se inicia el cambio en el uso de la denominación y ya en 1990 nos indica que la evaluación debe ser usada para apoyar los procesos de aprendizaje, más que para medir el rendimiento actual o el del pasado más o menos inmediato.

La evaluación debe ser diseñada para capturar evidencias respecto de cómo se está produciendo el proceso de aprendizaje y proporcionar información a los agentes interesados

(básicamente estudiantes y profesores) de cómo deben orientar sus estrategias o estilos de enseñanza-aprendizaje para conseguir actuaciones más competentes.

En palabras de Glaser, necesitamos servirnos de un amplio espectro de pruebas de todo tipo para ser capaces de evaluar amplios aspectos de la red cognitiva del sujeto y superar los modelos centrados únicamente en la preocupación por obtener mediciones precisas y que únicamente nos proporcionan información respecto de la adquisición y retención de conocimientos aislados.

Raven nos señala la necesidad de desarrollar modelos que evalúen actuaciones relacionadas con objetivos ambiciosos y valiosos y no aquellos orientados a evaluar elementos cognitivos de carácter aislado. En definitiva, nos insta a reorientar nuestros esfuerzos y movernos fuera de lo meramente cognitivo y afrontar, evaluar y establecer su relación con los procesos metacognitivos.

Por su parte Goldstein, argumenta la necesidad de abandonar una concepción de la aplicación de pruebas (lo que denominaríamos en términos anglosajones como “testing”), entendida como una actividad estática que no provoca ninguna clase de efecto sobre los estudiantes. Preconiza un modelo más interactivo de evaluación bajo el cual no se asume que la habilidad de responder a los items de la prueba permanece inalterada a lo largo de la propia ejecución. Este posicionamiento, y en cierta manera los anteriores, preconiza la aparición de los movimientos de “evaluación auténtica”, bajo estos movimientos se desvanece la frontera entre actividades de evaluación y actividades de aprendizaje.

*Evaluar para certificar o evaluar para desarrollar*

Con todo, el cambio producido también genera sus contradicciones; así concretamente las pruebas construidas desde los principios alternativos, por las características de su propio diseño, difícilmente superan los principios de calidad de las pruebas basadas en la teoría

psicométrica, concretamente es muy difícil plantearse determinar su fiabilidad y situarlas en un contexto de estandarización.

Estas dos características son muy necesarias cuando se pretenden aplicar las pruebas con fines asociados a la petición de responsabilidades, en los que es muy importante, para la propia credibilidad de la prueba, garantizar la consistencia interna de los reactivos y la objetividad en sus condiciones de aplicación y corrección.

Las pruebas diseñadas y pensadas para ser pasadas a nivel nacional o internacional precisan, al margen de su validez, es decir, de su capacidad de capturar información sustantiva y válida para los propósitos específicos de la misma, recoger la información respecto de los sujetos o de las muestras de ellos, de forma que garantice su comparabilidad.

Fenómenos como el de la globalización implican la necesidad de diseñar las políticas educativas de los países y de las comunidades internacionales a partir de los datos obtenidos a partir de procesos evaluativos de carácter muy amplio. La información de cada país debe leerse y analizarse a la luz de un sinfín de indicadores, muchos de ellos obtenidos a partir de información evaluativa establecida mediante pruebas que para arrojar resultado que puedan ser comparables y permitan su interpretación para cada uno de los diferentes ámbitos, debe estar garantizado a través de su diseño y proceso de construcción que cumplen con los principios anteriormente señalados, de lo contrario perderían absolutamente toda credibilidad.

Con ello se nos plantea el gran dilema: afrontamos un nuevo enfoque de la evaluación basado en la generación de pruebas que nos permitan profundizar en la comprensión de la producción de la cognición y en los procesos de aprendizaje de los sujetos de forma que seamos capaces de desarrollar habilidades y competencias de carácter superior pensando en cada sujeto como único e intransferible y renunciando a todo proceso generalizador, o, por el contrario, aceptamos las crecientes presiones por enfocar los procesos evaluativos en la obtención de grandes masas de datos provenientes de ampliar muestras extraídas de forma aleatoria y de acuerdo al cumplimiento estricto de rígidas condiciones psicométricas específicas para la



obtención de puntuaciones que por sus características puedan ser utilizadas en el diagnóstico de los sistemas educativos, en la generación de políticas educativas y en la certificación de resultados. En definitiva, evaluar para el desarrollo del sujeto o evaluar para la certificación de situaciones.

Nuestra personal posición es que es posible compatibilizar ambos enfoques. Si alguna característica tiene la evaluación moderna es la necesidad de utilizar una pluralidad de enfoques y el uso estratégico y combinado de diferentes aproximaciones evaluativas.

La información evaluativa que proporcionan ambos paradigmas evaluativos permiten delimitar con mayor precisión los contextos a evaluar, cada uno proporciona información situada en niveles distintos pero que están interaccionados.

Sin embargo, el uso combinado de dos enfoques no quiere decir mixtificado, es decir, la planificación se establece desde supuestos distintos y el desarrollo se plantea con técnicas diferenciadas. Simplemente es en el marco de la realidad que se analiza donde cada enfoque arroja diferentes tipos de luz que ayudan a comprender mejor los espacios de evaluación y decisión.

#### *Hacia una nueva teoría de la evaluación educativa*

Evidentemente se precisa una mayor reflexión en torno a la evaluación educativa. Debemos superar la tendencia natural a pensar en términos simplistas y creer que una aproximación evaluativa es mejor que la otra. El camino para por establecer una interacción dinámica y creativa entre las dos formas de abordar la evaluación que nos aporte una mayor comprensión de los ámbitos a evaluar.

Desde esa visión comprensiva e interactiva es necesario desarrollar una amplia metodología de evaluación basada en una u otra aproximación pero dejando muy claro cómo debe producirse su uso adecuado. Es importante clarificar como, cuando y donde debe usarse para producir información sustantiva que clarifique los contextos evaluativos y no los contamine.

El problema que posiblemente hemos tenido hasta ahora es el del uso indiscriminado y poco refinado de las metodologías evaluativas. No debe usarse una actividad pensada desde una óptica para resolver problemas para los que nos fue diseñada. Usar complementariamente las metodologías implica controlar su uso desde las necesidades de la propia realidad. La realidad formula sus preguntas y hay que ver si las respuestas que se generan desde una posición son las adecuadas y a qué nivel del planteamiento del problema puede usarse aquella información.

### *Diseño de pruebas*

La construcción de instrumentos de medida, “testing”, constituirá uno de los apartados más importantes dentro del texto. Nos plantearemos las características que deben reunir tanto para atender a las condiciones de calidad propias de la teoría de los tests (fiabilidad y validez), como a la substantividad y relevancia de la información evaluativa que nos proporcionan los reactivos elaborados.

## **1.14 Estructura del libro**

Consecuentemente, el texto contemplará dos grandes apartados uno dedicado a la medición, es decir, a los planteamientos evaluativos basados en la psicometría y en la teoría de la medición y el otro orientado a la evaluación educativa, más centrado en los denominados procesos de "assessment".

El libro consta de diez capítulos, en este primero hemos presentado la materia, estableciendo las bases de su evolución. Mantenemos intacto para la disciplina el nombre de Medición y Evaluación Educativa, entendiendo que se trata de la misma disciplina, sólo que el énfasis ha pasado, tal como queda descrito en este primer capítulo introductorio, de la medición a la evaluación.

En el segundo capítulo abordamos los elementos más básicos de la Teoría de la Medición, muy especialmente en lo que hace referencia a las escalas de medida y al tratamiento de los datos desde las perspectivas cuantitativa y cualitativa.

El tercer capítulo se inicia con la presentación de la Teoría Clásica de los Tests y de las características métricas que debe reunir una prueba. Evidentemente también se analizan las críticas a la teoría y se presentan las alternativas más importantes y conocidas.

El cuarto capítulo tiene un carácter eminentemente práctico. En él presentamos los procesos de elaboración de los instrumentos de medición y evaluación más usados en Educación. Por su tamaño se trata de un capítulo doble, pero hemos preferido mantenerlo único para no dañar la comprensión global del mismo.

El quinto capítulo se introduce en la naturaleza y conceptualización de la evaluación. Se abordan sus funciones y se caracteriza su carácter procesual, presentando las fases a seguir en su aplicación. También se estudian los ámbitos más significativos de la realidad educativa que deberían abordarse en la práctica evaluativa.

En el sexto capítulo nos centramos, de entre todos los ámbitos presentados en el capítulo anterior, en la evaluación de los aprendizajes de los alumnos, tanto desde la perspectiva tradicional como desde los nuevos planteamientos alternativos. En este capítulo se producen algunas redundancias con el capítulo cuatro, sin embargo hemos mantenido íntegramente ambas exposiciones, porque cualquier recorte podría eliminar elementos básicos para la comprensión global del discurso.

### **1.15 Análisis de datos en la medición educativa**

A nuestro entender, una manera adecuada de descubrir las características específicas del análisis de los datos es situarlo en el contexto general de la evaluación de programas y, a partir de ese punto, comenzar la búsqueda de las notas distintivas que lo caracterizan.

Y, por extraño que pueda parecer, hay una cuestión en la que todos los autores básicamente coinciden, y es en el hecho de considerar la evaluación de programas como una actividad investigadora.

En consecuencia, si asumimos lo antedicho, al análisis de datos le correspondería la tarea que habitualmente se cita cuando se habla de esta fase del proceso de investigación empírica en ciencias sociales y humanas. A saber: proporcionar evidencia suficiente o respuesta a las hipótesis y preguntas planteadas mediante técnicas analíticas (habitualmente estadísticas en el paradigma cuantitativo) transformando los datos en resultados.

Pero evidentemente la cuestión no es ni mucho menos tan sencilla. La evaluación de programas es una actividad investigadora, pero evaluación e investigación no son términos intercambiables. Un primer matiz es el de considerar la evaluación de programas como un caso típico de investigación aplicada hablándose también de investigación evaluativa.

Dado que la evaluación de programas es una actividad investigadora, la consecuencia que de ello se deriva es que las fases de la misma no han de diferir excesivamente de las propias del proceso investigador.

Y de hecho así ocurre. Si examinamos nos daremos cuenta que casi siempre, después de las decisiones sobre el diseño y de la etapa de recogida de información, aparece la fase del análisis de los datos, a la que suele seguir la elaboración del informe y la puesta en práctica de las conclusiones.

Y, por descontado, nada hay que objetar a esta secuencia en la medida en que refleja el orden lógico del proceso investigador. Los matices suelen surgir en el número de etapas que se distinguen, o en las diferencias según los enfoques de la evaluación.

Pero desde nuestra particular perspectiva centrada en la fase del análisis de los datos, nos interesa señalar que, en primer lugar, las decisiones sobre qué técnicas emplear para analizar

los datos van a estar a su vez, lógicamente en función de las decisiones que en las fases previas se hayan ido tomando: qué enfoque se ha dado a la evaluación (cuantitativo/cualitativo/mixto), qué preguntas o cuestiones se plantean, qué diseños se han empleado, qué unidades se han definido, cómo han sido seleccionadas y —si ha sido el caso— asignadas, qué tipo de información y mediante qué instrumentos se ha recogido.

Segundo, si bien esta secuencia es lógicamente consistente, no debe ser tomada al pie de la letra en el sentido de que después de recoger la información, ésta se organiza y se analizando respuesta a las preguntas planteadas. Evidentemente el proceso no es ni mucho menos tan lineal porque el análisis de los datos es un proceso abierto, cíclico e interactivo: los resultados de los análisis preliminares condicionan los siguientes y así sucesivamente, a menudo modificando las previsiones iniciales.

Este carácter abierto de la fase del análisis es válido y aplicable a cualquier proceso de investigación, aunque lógicamente en mayor grado cuanto más exploratorio sea el enfoque adoptado. Pero en el caso de la evaluación de programas cobra especial relevancia habida cuenta del papel que los agentes implicados o interesados pueden jugar en esta fase. Cuanto más abierto y participativo sea el enfoque, más interactivo y cíclico es el análisis de los datos.

Pero además de esta ordenación lógica basada en las fases del proceso investigador, también es posible distinguir etapas en función del desarrollo, de la secuencia del propio programa a evaluar. Aquí, en función de la fase del proceso en que se centre la actividad evaluadora, nos podemos encontrar con distintos tipos de evaluación: evaluación de necesidades; evaluación del diseño; evaluación de la evaluabilidad; evaluación de la implementación; evaluación de la cobertura; evaluación del proceso; evaluación de resultados, impactos o producto; y, evaluación económica.

Desde nuestra habitual perspectiva centrada en el análisis de los datos, en primer lugar señalemos que el análisis de datos encuentra su lugar, digamos, natural en la fase o contexto de la evaluación de resultados o de impactos.

Segundo, que al igual que antes, las técnicas analíticas a emplear estarán también en función de los resultados y de las consiguientes decisiones que se hayan ido tomando en evaluaciones previamente realizadas (del proceso, por ejemplo).

Tercero, que las técnicas analíticas pueden emplearse —y de hecho se emplean— en muchas fases o tipos de evaluación, y no sólo en la evaluación de resultados o de impactos a pesar de que algunos consideren ésta la fase analítica por excelencia.

Cualquier tipo de evaluación requiere el procesamiento de la información recogida. La gran cantidad de datos brutos o directos que se suelen obtener en las evaluaciones dificulta su interpretación, y por este motivo se destina a continuación un apartado sobre el análisis de los datos, que consiste básicamente en reducir y sintetizar la información de forma que sea más comprensible.

Dado el elevado número de pruebas y procedimientos existentes para realizar el análisis de datos, ya sean éstos cuantitativos o cualitativos, este apartado se centrará en los procesos más habituales de la investigación educativa y sobre todo en aquellos más utilizados para la elaboración o uso de los instrumentos de medición y evaluación educativa.

## **1.16 Escalas de medidas**

Una de las definiciones ampliamente aceptada del concepto "medida" es la asignación de números a objetos o acontecimientos de acuerdo a unas reglas que le dan significado. Para asignar números, o mejor dicho en este caso "numerales", Stevens (1946, 1959) distinguió cuatro tipos de reglas que dieron lugar a las ya clásicas cuatro escalas de medida jerarquizadas: nominal, ordinal, de intervalo y de razón. En 1957, el propio Stevens añadió la escala logarítmica.

Stevens mantiene el supuesto de isomorfismo entre las propiedades del sistema formal o numérico y del sistema empírico. Considera que las propiedades de las escalas contienen la

información suficiente para representar las propiedades del mundo real en virtud del isomorfismo entre ambos sistemas. Para poder medir un objeto debe existir una relación isomórfica. Esto significa que debe haber igualdad formal entre las operaciones que se pueden hacer con los numerales y las que se pueden realizar con los objetos medidos. Pero el requisito de establecer el isomorfismo que exige la Física no se cumple siempre en el ámbito socio-educativo (Del Rincón et al., 1995). La solución propuesta es admitir que en ciencias sociales son posibles diversos niveles de medición y, por lo tanto, que existen diversas operaciones formales que se pueden realizar con las mediciones.

### **1.17 El análisis cuantitativo**

Los procesos aplicados en el análisis cuantitativo de los datos son estadísticos y se resumen en las siguientes fases:

- Codificación de los datos (cuando están medidas en escala nominal u ordinal).
- Creación de una matriz con los datos codificados.
- Selección de la prueba estadística a aplicar en función del tipo de variable, el número y tamaño de las muestras, el número de variables y el objetivo del análisis describir, comparar, asociar.
- Aplicación de la prueba o fórmula estadística.
- Interpretación del resultado.

Para realizar las pruebas estadísticas hoy día se utilizan programas informáticos que facilitan la gestión de los datos y los cálculos matemáticos. Entre los más difundidos en el ámbito de las ciencias sociales están el SPSS (Statistical Package for the Social Science), el BMDP (BioMeDical Package) y el SAS (Statistic Analysis System), entre otros. El SAS es uno de los más completos y perfeccionados pero tiene poca difusión en el ámbito educativo. El BMDP tiende a ser más utilizado en el ámbito biomédico. El SPSS permite realizar la casi totalidad de pruebas estadísticas necesarias en el ámbito educativo, aunque no se suele adquirir a nivel personal o de pequeñas organizaciones debido a su coste económico. En cambio, la enorme difusión que

están teniendo las hojas de cálculo (tipo Excel de Microsoft, por ejemplo) y su potencial estadístico en continuo crecimiento permiten realizar algunas pruebas estadísticas interesantes a profesionales independientes e incluso a organizaciones como centros escolares. Los fundamentos de las pruebas estadísticas se hallan en múltiples obras publicadas, de las cuales, como ejemplo, pueden ser Cuadras et al. (1988) y Christensen (1990).

### **1.18 Interpretación normativa de puntuaciones**

Interpretar la puntuación de un test correctamente es básico para tomar decisiones en el ámbito de la selección de personas.

Cuando un individuo realiza un test psicométrico, proporciona una puntuación directa sobre la base de sus respuestas a los elementos de la prueba.

En sí mismas, esas puntuaciones directas carecen de significado. De hecho, el que una persona obtenga un puntaje de 40 en una prueba de razonamiento abstracto no nos dice gran cosa sobre su aptitud general. ¿Son fáciles las preguntas a las que ha respondido? ¿Son difíciles? ¿Cómo hubieran respondido otras personas?

Interpretar la puntuación de un test solo es posible atendiendo a sus “normas”. Esto incluye su muestra de tipificación que es el grupo representativo con el que hay que establecer comparaciones. De esta forma, la puntuación directa se contrasta con dicha muestra y, entonces, sí es posible estimar si el puntaje del sujeto coincide con el promedio de la muestra, es superior o inferior y en qué grado se desvía de ese promedio.

Así, se establece la posición relativa del individuo con referencia a su grupo normativo, a la muestra de tipificación. Pero para la correcta interpretación de los resultados de un test es necesario transformar las puntuaciones directas en medidas relativas, lo que permitirá conocer la situación del individuo respecto al grupo normativo.



Con el objetivo de introducir cómo interpretar la puntuación de un test, es previo destinar unas líneas a las diferencias individuales y la forma de su distribución.

Así, consideremos que todas las personas no son iguales. Esto lo vemos todos los días. De hecho, no todas tienen la misma estatura, ni la misma capacidad para recordar nombres o resolver problemas matemáticos.

Cualquier característica o rasgo, al medirlo se distribuye de un modo particular. Además, lo hace de una manera muy semejante, siempre que se seleccionen individuos al azar y que sean representativos de la población en general.

### **1.19 El análisis cuantitativo**

Las técnicas cuantitativas utilizan números para poder evaluar la información, estos números son analizados a través de análisis estadísticos para comprobar o evaluar hipótesis o procesos, a la vez dar un seguimiento y verificar el progreso que se va generando.

La recolección de datos cuantitativos se basa en:

- La utilización de instrumentos que son confiables y han sido validados
- Los datos obtenidos a través de estos instrumentos son objetivos y precisos y son analizados utilizando técnicas estadísticas.
- Están enfocados hacia datos que son medibles y observables.

### **1.20 Teorías sobre las técnicas de obtención de información**

Generalmente se distinguen dos tipos de técnicas: cualitativas y cuantitativas, y se han clasificado así en atención del predominio interno de las características para aplicarlas y procesar la información que se obtiene; ya contengan en ello números o no; aunque la clasificación es endeble porque en ambos tipos de técnicas es frecuente que se utilicen números y elementos discursivos e interpretativos.

Una clasificación más responde a criterios de aplicación; así, se habla de técnicas de investigación documental, que se orientan a trabajo con información contenida en soportes documentales como impresos, grabaciones o relacionada con computadoras y redes virtuales; o técnicas de campo, cuando se trabaja con personas, con lugares o con aparatos.

Algunas de las técnicas cualitativas más utilizadas son las siguientes:

- Técnicas de investigación documental (procedimientos orientados a la aproximación, procesamiento y recuperación de información contenida en documentos, independientemente del soporte documental en que se hallen) Ejemplos de estas técnicas son el aparato crítico (conjunto de apoyos al texto principal de un documento científico, según normas internacionales), las técnicas de lectura (aproximación al documento) y la reseña, el resumen etc. (recuperación de la información).
- Técnicas para la obtención de información en campo (procedimientos del tipo sujeto-objeto como la observación directa, o sujeto-sujeto como la entrevista) que se aplican a procesos sociales u objetos.
- Técnicas de procesamiento de información (procedimientos que se aplican a la información obtenida en campo y cuya base es la interpretación, por ejemplo, la TKJ o la Delphi).

## 1.21 Teoría clásica de Tests

Los tests constituyen seguramente la tecnología más sofisticada de la que disponen los psicólogos para ejercer su profesión, por eso no es infrecuente que la sociedad identifique a los psicólogos con los tests. Naturalmente, unos psicólogos utilizan los tests más que otros, dependiendo de su campo profesional y de su forma de trabajar. Los tests son muestras de conducta que permiten llevar a cabo inferencias relevantes sobre la conducta de las personas. Bien utilizados son herramientas claves en la profesión del psicólogo. No conviene olvidar que los tests nacen con un afán de objetividad y justicia, para evaluar a las personas por lo que realmente valen, evitando evaluaciones sesgadas por aspectos tales como la cuna, la clase social,

la raza, el sexo, las creencias, las cartas de recomendación, y otros sistemas de evaluación subjetivos. Unas veces estos nobles fines se han alcanzado mejor que otras, pero ésa era y sigue siendo la idea central, evaluar a todos por el mismo rasero.

El enfoque clásico es el predominante en la construcción y análisis de los tests, así, por ejemplo, los diez tests más utilizados por los psicólogos españoles citados en el apartado anterior, todos ellos, sin excepción, han sido desarrollados bajo la óptica clásica. Sólo este dato ya deja bien patente la necesidad de que los profesionales entiendan perfectamente la lógica clásica, sus posibilidades y sus limitaciones. Antes de entrar en la lógica de la teoría clásica, hay que señalar que hincan sus raíces en los trabajos pioneros de Spearman de principios del siglo XX (Spearman, 1904, 1907, 1913). Lleva por lo tanto unos cien años en el circuito, así que se ha ganado por méritos propios el adjetivo de clásica. A partir de esos años se produce un rápido desarrollo y para 1950 lo esencial ya está hecho, así que Gulliksen (1950) lleva a cabo la síntesis canónica de este enfoque. Más adelante serán Lord y Novick (1968) quienes lleven a cabo una reformulación de la teoría clásica y abran paso al nuevo enfoque de la TRI que veremos luego.

## **1.22 Críticas y alternativas a la teoría clásica de Tests**

Del enfoque de la teoría clásica bien podría decirse que goza de muy buena salud, hay pocas dudas de su utilidad y eficacia, baste decir, por ejemplo, que la gran mayoría de los tests editados en España, prácticamente todos, están desarrollados y analizados dentro de este marco. Ahora bien, si es así, la pregunta obligada es por qué hacen falta otras teorías de los tests, o, en otras palabras, ¿qué problemas de medición no quedaban bien resueltos dentro del marco clásico para que se propongan nuevas teorías? Pues bien, había dos cuestiones básicas que no encontraban buena solución en la teoría clásica y que hacían que la medición psicológica no fuese homologable a la que exhibían otras ciencias empíricas.

Veamos la primera: dentro del marco clásico, las mediciones no resultan invariantes respecto al instrumento utilizado. Se preguntarán con razón qué quiere decir exactamente esa afirmación un tanto críptica. Es muy sencillo, si un psicólogo evalúa la inteligencia de tres

personas distintas con un test diferente para cada persona, los resultados no son comparables, no podemos decir en sentido estricto qué persona es más inteligente. Esto es así porque los resultados de los tres tests no están en la misma escala, cada test tiene la suya propia. Esto puede sorprender a los psicólogos usuarios habituales de la teoría clásica, acostumbrados en la práctica a comparar la inteligencia de personas que han sido evaluadas con distintos tests de inteligencia. Para hacerlo se transforman las puntuaciones directas de los tests en otras baremadas, por ejemplo en percentiles, con lo que se considera que se pueden ya comparar, y de hecho así se hace. Este proceder clásico para solventar el problema de la invarianza no es que sea incorrecto, pero, amén de poco elegante científicamente, descansa sobre un pilar muy frágil, a saber, se asume que los grupos normativos en los que se elaboraron los baremos de los distintos tests son equiparables, lo cual es difícil de garantizar en la práctica. Si eso falla la comparación se viene abajo. No hay duda que lo más deseable científicamente sería que los resultados obtenidos al utilizar distintos instrumentos estuviesen en la misma escala, y todo quedaría resuelto de un plumazo, pues bien, por extraño y contra intuitivo que parezca eso es precisamente lo que va a conseguir el enfoque de la TRI. Este nuevo enfoque de la TRI va a suponer un gran avance para la medición psicológica, propiciando un gran desarrollo de nuevos conceptos y herramientas psicométricas.

La segunda gran cuestión no bien resuelta dentro del marco clásico era la ausencia de invarianza de las propiedades de los tests respecto de las personas utilizadas para estimarlas. En otras palabras, propiedades psicométricas importantes de los tests, tales como la dificultad de los ítems, o la fiabilidad del test, estaban en función del tipo de personas utilizadas para calcularlas, lo cual resulta inadmisibile desde el punto de vista de una medición rigurosa. Por ejemplo, la dificultad de los ítems, o los coeficientes de fiabilidad dependen en gran medida del tipo de muestra utilizada para calcularlos. Este problema también encontrará una solución adecuada dentro del marco de la TRI.

Aparte de estas dos grandes cuestiones, había otras menores de carácter más técnico a las que la teoría clásica no daba una buena solución. Por ejemplo, cuando se ofrece un coeficiente de fiabilidad de un test en el marco clásico, como el coeficiente alfa de Cronbach (1951), se está

presuponiendo que ese test mide con una fiabilidad determinada a todas las personas evaluadas con el test, cuando tenemos evidencia empírica más que suficiente de que los tests no miden con la misma precisión a todas las personas, dependiendo la precisión en gran medida del nivel de la persona en la variable medida. El nuevo marco de la TRI va a solucionar este problema ofreciendo la Función de Información, que permite estimar la fiabilidad de la prueba en función del nivel de la persona en la variable medida.

Además de estas cuestiones centrales, la TRI va a generar toda una tecnología psicométrica nueva que cambiará para siempre la forma de hacer psicometría; véase por ejemplo en este mismo número monográfico el trabajo de Olea, Abad y Barrada (2010). Ahora bien, conviene dejar muy claro que estos nuevos modelos de TRI de ninguna manera invalidan el enfoque clásico, si bien constituyen un excelente complemento que en determinadas circunstancias dan solución a problemas mal resueltos en el marco clásico. Ambas tecnologías conviven perfectamente en la construcción y análisis de los tests, igual que coches y aviones lo hacen en el transporte, valga la analogía, unos son aconsejables en determinadas situaciones, y otros lo son en otras.

## UNIDAD II

### ELABORACION DE INSTRUMENTOS DE MEDICION Y EVALUACION EDUCATIVA

#### 2.21 Clasificación de los instrumentos de medición y evaluación educativa

Esta clasificación atiende a diferentes criterios. Por tanto, se emplean uno u otro en función del propósito de la evaluación, a los impulsores o ejecutores de la misma, a cada situación concreta, a los recursos con los que contamos, a los destinatarios del informe evaluador y a otros factores.

Según su finalidad y función

a) Función formativa: la evaluación se utiliza preferentemente como estrategia de mejora y para ajustar sobre la marcha, los procesos educativos de cara a conseguir las metas u objetivos previstos. Es la más apropiada para la evaluación de procesos, aunque también es formativa la evaluación de productos educativos, siempre que sus resultados se empleen para la mejor de los mismos. Suele identificarse con la evaluación continua.

b) Función sumativa: suele aplicarse más en la evaluación de productos, es decir, de procesos terminados, con realizaciones precisas y valorables. Con la evaluación no se pretende modificar, ajustar o mejorar el objeto de la evaluación, sino simplemente determinar su valía, en función del empleo que se desea hacer del mismo posteriormente.

Según su extensión

a) Evaluación global: se pretende abarcar todos los componentes o dimensiones del alumnos, del centro educativo, del programa, etc. Se considera el objeto de la evaluación de un modo holístico, como una totalidad interactuante, en la que cualquier modificación en uno de sus componentes o dimensiones tiene consecuencias en el resto. Con este tipo de evaluación, la comprensión de la realidad evaluada aumenta, pero no siempre es necesaria o posible. El modelo más conocido es el CIPP de Stufflebeam.

b) Evaluación parcial: pretende el estudio o valoración de determinados componentes o dimensiones de un centro, de un programa educativo, de rendimiento de un alumnos, etc.

Según los agentes evaluadores

a) Evaluación interna: es aquella que es llevada a cabo y promovida por los propios integrantes de un centro, un programa educativo, etc. A su vez, la evaluación interna ofrece diversas alternativas de realización: autoevaluación, heteroevaluación y coevaluación.

\* Autoevaluación: los evaluadores evalúan su propio trabajo (un alumno su rendimiento, un centro o programa su propio funcionamiento, etc). Los roles de evaluador y evaluado coinciden en las mismas personas.

\* Heteroevaluación: evalúan una actividad, objeto o producto, evaluadores distintos a las personas evaluadas (el Consejo Escolar al Claustro de profesores, un profesor a sus alumnos, etc.)

\* Coevaluación: es aquella en la que unos sujetos o grupos se evalúan mutuamente (alumnos y profesores mutuamente, unos y otros equipos docentes, el equipo directivo al Consejo Escolar y viceversa). Evaluadores y evaluados intercambian su papel alternativamente.

b) Evaluación externa: se da cuando agentes no integrantes de un centro escolar o de un programa evalúan su funcionamiento. Suele ser el caso de la "evaluación de expertos". Estos evaluadores pueden ser inspectores de evaluación, miembros de la Administración, investigadores, equipos de apoyo a la escuela, etc.

## 2.22 Pruebas de rendimiento de ensayos

El objetivo de las experiencias comentadas era doble, por un lado pretendíamos hacer patente la difícil objetividad de las pruebas de ensayo pero también queríamos evidenciar, en caso de utilizarlas, la posibilidad que tenemos para reducir al máximo la subjetividad.

Los resultados del estudio avalan de forma clara las dos previsiones. Por un lado queda patente la poca concordancia entre evaluadores cuando se corrigen las pruebas de ensayo y, por otro, vemos como a medida que precisamos los criterios de evaluación la subjetividad disminuye. Tanto con profesores como con estudiantes, la dispersión máxima de notas se produce en la

primera situación, ésta se reduce en la segunda evaluación, cuando se establecen ciertos criterios de evaluación, y la dispersión es mucho menor si se establecen puntuaciones concretas para cada uno de dichos criterios.

De los resultados de nuestro estudio se desprende un hecho que no estaba expresamente buscado. Las diferencias encontradas entre las dos poblaciones de evaluadores podrían ser atribuidas a un posible efecto “halo”, otro de los grandes inconvenientes que presentan este tipo de pruebas. En la primera evaluación sin criterios (S-I), vemos que los profesores asignaron menos notas inferiores a cinco que los estudiantes ( 2 % vs 12 % ). Seguramente este hecho se debería a que en el caso de los profesores, éstos evaluaron el examen creyendo que era el de un compañero que compartía el taller. Por el contrario, en el caso de los estudiantes se evaluaba un examen de una persona anónima. Este efecto aún se mantuvo en la segunda evaluación cuando ya se incluían criterios concretos (25 % vs 45 %) pero desapareció en la tercera evaluación cuando la precisión para puntuar era máxima (76 % vs 75 %).

### **2.23 Pruebas objetivas de rendimiento de test**

Las pruebas objetivas son instrumentos de medida, elaborados rigurosamente, que permiten evaluar conocimientos, capacidades, destrezas, rendimiento, aptitudes, actitudes, inteligencia, etc.

Las pruebas objetivas son un recurso utilizado para la evaluación diagnóstica, para la formativa y para la sumativa.

Se caracterizan por los puntos siguientes:

- Las respuestas tienen que ser breves, muy concretas, que no puedan dar lugar a cuestionarse su corrección o incorrección.
- Tener una única solución correcta.
- Favorecer la objetividad en la corrección.
- Las respuestas cerradas: el estudiante sólo tiene que escoger, señalar o completar respuestas con elementos muy concretos.



Dentro de las pruebas objetivas podemos diferenciar diferentes tipos de ítems en función de los objetivos y productos específicos del aprendizaje que queremos evaluar:

- Pruebas de respuesta breve
- Pruebas de completar
- Pruebas de discriminación
- Pruebas de ordenación
- Pruebas de elección múltiple
- Pruebas de asociación
- Pruebas de localización o identificación
- Pruebas multiítem

Para la elaboración de este tipo de pruebas hace falta invertir mucho tiempo para confeccionar y redactar las preguntas adecuadamente, ya que su formulación tiene que ajustarse a las características de los ítems objetivos.

La agilidad con que puede responderse a las preguntas tiene la ventaja de que pueden cubrir sobradamente el contenido del curso. Así, se recomiendan para medir conocimientos de hechos y conceptos.

En cambio, no permiten evaluar objetivos didácticos que impliquen aportación libre del estudiante, ya que eliminan toda posibilidad creativa porque parten de una respuesta totalmente prefijada.

Por lo tanto, no pueden medir la capacidad para seleccionar y organizar ideas, ni las habilidades de escribir, ni ciertos tipos de habilidades para resolver problemas.

Con respecto a su influencia sobre el aprendizaje, ayudan a los estudiantes a conseguir un conocimiento de hechos concretos y a desarrollar la capacidad de darse cuenta de sutiles diferencias entre éstos. No obstante, fomentan hábitos de estudio parciales y atomizados, por lo que tienen que combinarse con otros recursos didácticos y con otros tipos de evaluación.

En lo que concierne a la corrección y calificación:

- Permiten una corrección y calificación rápida. Pueden utilizarse medios mecánicos (plantillas, ordenadores...).
- Los resultados son independientes del estado de ánimo y la subjetividad del docente a la hora de corregir.
- La calificación objetiva es rápida, fácil y consistente. No influyen otros factores ajenos al contenido de las pruebas como la presentación, el estilo, etc.
- Los resultados pueden compararse ya que todos los estudiantes están evaluados con los mismos criterios.
- No permiten extraer consecuencias sobre el nivel general del grupo a no ser que se compruebe previamente el índice de dificultad de las pruebas.

## 2.24 Escalas de actitud

Si se van a estudiar las actitudes se debe contar con técnicas que las midan para comprender como se forman y cambian. Existen comúnmente varios métodos para estudiar las actitudes desde el punto de vista psicológico y obtener una medida cuantitativa. Estos comprenden el empleo de las escalas de ordenación, de intensidad, las escalas de actitudes desarrolladas por Bogardus, Dood, Thurstone, Crespi, Guttman y likert. También se utiliza el Diferencial Semántico y estrategias complementarias tales como: encuestas, cuestionarios, entrevistas así como la observación de la conducta y reacciones fisiológicas del organismo. De éstos métodos se describirán en un apartado siguiente las técnicas relevantes al estudio.

En la medición de la actitud se usa típicamente el término escale de actitud, donde en su aplicación se obtienen afinaciones positivas o negativas de una manera graduada en forma significativa y cuantitativa.

Según Cornejo (1990) se puede considerar una serie de características métricas en toda actitud

- a) la actitud posee una dirección negativa positiva
- b) posee un punto neutral de indefinición o indecisión (el punto O)
- e) posee intensidad o graduación
- y d) posee consistencia referida al grado de relación que guardan entre si los distintos componentes de una actitud (p.104).

Para cualquier escala de actitudes se deberán reunir una serie de ítems que indiquen el grado de afecto positivo o negativo que está asociado con el objeto psicológico o que existe hacia el objeto psicológico. Una vez que se tienen los ítems, se seguirán procedimientos específicos para cada tipo de escala y por su propio proceso de elaboración; según Cornejo (1990) "las escalas poseen altos coeficientes de confiabilidad y validez por su propio proceso de elaboración (p.105).

La medición de las actitudes por medio de encuestas de opinión y escalas se ha utilizado ampliamente para diversos fines, uno de los principales es proporcionar índices para medir los cambios en las actitudes de estudiantes relacionados con un curso determinado u otro programa educativo. La información resultante puede usarse para fines tales como estimar la eficacia de la educación y el adiestramiento.

La escala de Likert (también denominada método de evaluaciones sumarias) se denomina así por Rensis Likert, quién publicó en 1932 un informe donde describía su uso. Es una escala psicométrica comúnmente utilizada en cuestionarios, y es la escala de uso más amplio en encuestas para la investigación, principalmente en ciencias sociales. Al responder a una pregunta de un cuestionario elaborado con la técnica de Likert, se especifica el nivel de acuerdo o desacuerdo con una declaración (elemento, ítem o reactivo o pregunta).

La escala de Likert es un método de escala bipolar que mide tanto el grado positivo como neutral y negativo de cada enunciado.

La escala de Likert, al ser una escala que mide actitudes, es importante que pueda aceptar que las personas tienen actitudes favorables, desfavorables o neutras a las cosas y situaciones lo cual es perfectamente normal en términos de información. Debido a ello es importante considerar siempre que una escala de actitud puede y debe estar abierta a la posibilidad de aceptar opciones de respuesta neutrales.

Escalas Tipo Thurstone: El psicólogo norteamericano León I. Thurstone (1887 – 1955) quien consiguió cuantificar objetos propios del estudio de las ciencias sociales, prescindiendo de cualquier referencia a objetos físicos. Thurstone (1928) consideraba la opinión como la expresión verbal de la actitud. No podemos acceder directamente a la observación de la actitud,

pero la opinión verbal expresada de los sujetos puede servir como indicador de la actitud. Un índice de medida para la aceptación o rechazo de las opiniones de las personas permite obtener, indirectamente, una medida de sus actitudes, independiente de lo que esas personas "sientan en realidad" o de sus acciones.

El método de comparación por pares consiste en presentar a una muestra representativa de la población una lista previa de enunciados que, se piensa, pueden ser pertinentes para medir el grado de favorabilidad/desfavorabilidad de los individuos hacia cierto objeto. Sin embargo, los sujetos de esta muestra no actúan como sujetos que responden con su opinión a los enunciados o items, sino que hacen de jueces para discriminar la favorabilidad de esos enunciados. Es decir, no expresan su opinión sobre la afirmación que se les presenta, lo que se les pide es que manifiesten cuál es el grado de favorabilidad que expresan los enunciados con respecto al objeto de que tratan.

Hasta este punto, el método de comparación por pares y el de intervalos aparentemente iguales no difieren significativamente. Su diferencia esencial estriba en que en el método de comparación por pares los jueces tienen que comparar dos a dos todos los ítems que componen la escala indicando cuál de los dos es más favorable hacia el objeto de la actitud. Este método tiene la desventaja de que resulta enormemente costoso por la cantidad de comparaciones que deben realizar los jueces, sin embargo, cimienta los fundamentos teóricos de este tipo de escalas.

## **2.25 Técnicas sociométricas**

Los profesores tienen verdaderas dificultades para conocer la compleja red de relaciones que mantienen unidos entre sí a los diferentes alumnos de clase así como para descubrir las características sociales de los alumnos a quienes instruyen diariamente (las discrepancias entre lo que los docentes perciben y los resultados en las pruebas sociométricas incrementan a medida que aumenta la edad de los aprendices).

Según Moreno, las relaciones humanas en el seno de un grupo se desarrollan a dos niveles: uno de carácter afectivo y emocional y otro, selectivo y ordenado. El primer nivel constituye la estructura informal del grupo, mientras que el segundo da como resultado la estructura formal

del mismo. El test sociométrico es una técnica que permite la integración o puesta en relación de ambos niveles. El test sociométrico (sociograma) se define como un conjunto de procedimientos de observación y análisis de las relaciones que se establecen dentro de un grupo, procedimientos que se traducen en una serie de índices y representaciones gráficas que permiten evaluar y describir la estructura de las relaciones de carácter afectivo que existe en los pequeño grupos.

La tarea requerida en el test sociométrico es relativamente simple. Consiste en pedir a los miembros de un grupo que indiquen cuáles son los amigos/compañeros con los que preferirían realizar una determinada actividad así como aquéllos otros con quienes no les gustaría compartirla. Si bien la aplicación es sencilla, el análisis de la información obtenida no lo es tanto y puede requerir del docente un trabajo posterior de indagación y profundización (por ejemplo, mantener entrevistas con las figuras más representativas del test sociométrico).

Características del sociograma.

El test sociométrico presenta cuatro características. En primer lugar, se emplea en grupos naturales donde los miembros se conocen entre sí y los miembros que lo conforman son relativamente estables. En segundo lugar, permite la obtención de información sobre los deseos subjetivos de los miembros de un grupo. En tercer lugar, se realizan preguntas concretas que pueden condicionar la interpretación de los resultados. Por último, es importante tener en consideración las respuestas obtenidas, es decir, los resultados se utilizarán para iniciar una intervención (ofrecer soluciones reales para el conjunto del grupo).

Asimismo, el test sociométrico cumple con dos funciones esenciales. La primera de ellas es descriptiva donde se da cuenta de la estructura afectiva interna del grupo (estructura de atracciones y rechazos) permitiendo descubrir aspectos tales como: si el grupo está ya formado y presenta una organización interna en la que existen roles claramente diferenciados, cuáles son las personas más significativas del mismo por ser muy aceptados o rechazados, así como, vislumbrar algunas configuraciones sociométricas (el tipo de relaciones existentes entre dos o más personas en su seno).

La segunda función es explicativa y pretende relacionar ciertos aspectos de la estructura global del grupo con aspectos de su funcionamiento tales como: los valores y preferencias grupales (se supone que las personas más elegidas encarnan mejor los valores del grupo), los efectos que las condiciones externas tienen sobre la estructura grupal, la relación existente entre la estructura de sentimientos y la estructura de interacción y la relación entre solidaridad y obediencia a la norma grupal.

Posibilidades y limitaciones del sociograma.

Los resultados que se obtienen mediante la aplicación de un test sociométrico permiten entre otras cosas: tener una visión global de la estructura del grupo (por ejemplo, su grado de cohesión mediante el cálculo de algunos índices); conocer la posición relativa de cada uno de los miembros; conocer los subgrupos existentes (parejas, triángulos, subgrupos, reciprocidades,...); realizar un estudio más pormenorizado de las distintas figuras (líder, integrado, promedio, marginado, aislado, rechazado,...) en un intento de comprender las características que los definen; descubrir la existencia de posibles factores discriminatorios (religión, sexo, estatus socioeconómico,...); observar las variaciones de interés que se produzcan en el seno del grupo (aplicación del sociograma en dos momentos temporales diferentes); conocer la estructura informal del grupo y; por último; extraer conclusiones para mejorar las relaciones entre los miembros del grupo.

Atendiendo a las limitaciones es conveniente indicar las siguientes: la interpretación de los datos que refleja un sociograma exige una preparación amplia por parte del profesor u observador; los datos obtenidos son solo indicios que requieren ser corroborados mediante el empleo de otras técnicas (observación directa, entrevista,...) en un intento de conocer las posibles causas de por qué alguien es aceptado o rechazado; la estabilidad de los datos del sociograma es muy limitada (los cambios en las relaciones y sentimientos interpersonales suceden con mucha facilidad durante la etapa de Educación Infantil y Primaria) y; la limitación de las posibles elecciones (en ocasiones, se preferiría escoger a muchos más compañeros o compañeras).

## 2.26 Cuestionarios

Los cuestionarios, en el ámbito de la enseñanza, vienen a ser exámenes en formato test. Sirven para evaluar los conocimientos previos sobre una unidad didáctica determinada, planteando los ítems en un formato altamente estándar y siendo su respuesta única. Como el alumno no tiene que redactar, sino contestar a unos ítems de forma afirmativa o negativa, resulta fácil para al profesor corregir estas pruebas.

## 2.27 Entrevistas

Es un diálogo que se establece entre dos personas en el que una de ellas propone una serie de preguntas a la otra a partir de un guion previo. Se realiza con el fin de que el público pueda conocer la información de su persona, de su experiencia o conocimientos.

Tipos de entrevista: hay dos tipos de entrevista según su finalidad:

- a. Entrevista informativa: es aquella en la que el entrevistador pretende obtener información sobre un tema determinado.
- b. Entrevista psicológica: es aquella en la que se pretende conocer la personalidad o el ambiente que rodea al entrevistado.

## 2.28 Instrumentos de observación sistemática

Instrumento útil para informarnos sobre las motivaciones, intereses, progresos que los alumnos manifiestan en sus participaciones en clase.

En la observación sistemática, también llamada estructurada, se dispone de antemano, tanto el campo a estudiar (lugares y sujetos) como de los aspectos concretos o conductas sobre las que se va a centrar la atención.

El investigador suele establecer previamente una serie de categorías de observación (aspectos referidos a conductas, acciones, formas de respuesta, etc.) a partir de las cuales realizar la investigación.

El empleo de categorías de observación permite la cuantificación de las conductas observadas y su frecuencia entre otros datos.

Para registrar las categorías la observación sistemática suele hacer uso de muestreos. Los más frecuentes son:

- Muestreo de acontecimiento. Cuando la observación está orientada por la concurrencia de una serie de eventos determinados la selección toma forma cuando la acción pretendida tiene lugar, registrando en ella las variables definidas. Realizaremos por lo tanto tantas observaciones como veces acontece el evento escogido para el trabajo de campo.
- Muestreo de tiempo. Cuando la observación no se halla guiada por un acontecimiento determinado sino que asumimos el marco temporal del trabajo de campo como un continuo, el muestreo se orienta hacia la observación en intervalos de tiempo predefinidos, registrando en ese momento las acciones que se desarrollan en el marco de la observación de acuerdo con las variables relevantes.

## **2.29 Traducción de los instrumentos de medición y evaluación educativa**

En las últimas dos décadas, la práctica de traducir y adaptar instrumentos de medición educativa a otras lenguas o para culturas diferentes se ha hecho más frecuente a consecuencia de una tendencia hacia la economía global. Los resultados de comparaciones internacionales como TIMSS (por sus siglas en inglés, Estudio Sobre las Tendencias en Ciencias y Matemáticas) y PISA (por sus siglas en inglés, Programa Internacional para la Evaluación de Estudiantes) influyen cada vez más profundamente en la opinión pública y en las políticas educativas de los países participantes, cuyo desempeño en relación con el de otros países es un indicador de progreso académico.

Una consecuencia del auge actual de las comparaciones internacionales ha sido el desarrollo de nuevos y más sofisticados procedimientos de traducción de pruebas que tienen el fin de asegurar que los ítems de una prueba sean equivalentes en múltiples idiomas, a pesar de que existan importantes diferencias culturales entre los países participantes. Un componente clave en las comparaciones internacionales es el uso de un conjunto de lineamientos que garanticen una consistencia mínima en los procedimientos empleados por los diversos países. Por ejemplo, las naciones participantes en TIMSS-1995 emplearon los lineamientos para la traducción de



pruebas elaborados por Hambleton (1994) por encargo de la ITC (por sus siglas en inglés, Consejo Internacional de Evaluación) y, a partir de 2007, un nuevo conjunto de lineamientos entrará en vigor (ver Hambleton, 2005).

Otro componente clave en las comparaciones internacionales es el uso de un sistema de revisión de las traducciones de pruebas. Dicho sistema está basado en el empleo de un equipo de revisión de traducciones en el que no participan los traductores de los países participantes. Las traducciones de TIMSS-1995 fueron certificadas por una tercera instancia antes de se autorizara su aplicación por los países participantes (ver, por ejemplo, O'Connor y Malak, 2000).

Este artículo está dirigido a los profesionales involucrados en la traducción de pruebas en comparaciones internacionales. En él mostramos evidencia de que, aunque necesarios, los lineamientos para la traducción de pruebas y para la certificación de calidad de traducción pueden ser insuficientes para asegurar una adecuada traducción de los instrumentos. La implicación fundamental de este trabajo es que, para poder beneficiarse de su participación en comparaciones internacionales, los países deben asegurarse de tener procedimientos internos de revisión, que permitan implementar los lineamientos de traducción de una manera que sea sensible a aspectos sutiles pero muy importantes del uso del lenguaje en cada país.

El artículo está dividido en cuatro secciones. En la primera sección describimos un modelo conceptual para la revisión de pruebas. Dicho modelo incluye una variedad más amplia de aspectos de traducción que los que normalmente incluyen otros procedimientos. Estos aspectos se refieren a la producción de pruebas, la representación curricular y los aspectos sociales del uso del idioma. En la segunda sección describimos nuestro procedimiento de codificación de errores de traducción, basado en la facilitación de discusiones de comités interdisciplinarios de especialistas. En la tercera sección presentamos evidencia empírica de la sensibilidad de nuestro modelo de revisión. Esta evidencia proviene de nuestro análisis de la traducción mexicana de TIMSS-1995 y de un proyecto, actualmente en curso, en el que analizamos la calidad de la traducción mexicana de PISA-2003. Concluimos con un conjunto de recomendaciones para países que participan en comparaciones internacionales.

Marco conceptual para la revisión de traducción de pruebas

El desarrollo de nuestro marco conceptual se basó en el uso combinado de documentos normativos para la traducción de pruebas desarrolladas por el TIMSS (Hambleton, 1994; van de Vijver y Poortinga, 1996), procedimientos para la verificación de la calidad de la traducción de pruebas TIMSS (Mullis, Kelly y Haley, 1996), criterios usados por la American Translators Association (2003), normas y criterios que PISA emplea para determinar la adecuación cultural de los ítems de una prueba (Grisay, 2002; Maxwell, 1996), así como evidencia proveniente de nuestra propia investigación sobre el efecto de las características morfosintácticas de los ítems y los factores sociolingüísticos y epistemológicas, en las interpretaciones que hacen los estudiantes de los ítems en las áreas de ciencias naturales y matemáticas (Solano-Flores y Nelson-Barber, 2001; Solano-Flores y Trumbull, 2003; Solano-Flores, Trumbull y Kwon, 2003).

Nuestro marco conceptual está basado en cinco premisas: (a) inevitabilidad del error en la traducción de pruebas, (b) dimensiones del error de traducción, (c) relatividad de las dimensiones de error, (d) multidimensionalidad de los errores de traducción y (e) naturaleza probabilística de la aceptabilidad de la traducción de un ítem. A continuación explicamos cada una de estas premisas.

#### Inevitabilidad del error en la traducción de pruebas

En teoría, la equivalencia de constructo en dos idiomas es imposible de lograr, debido a que cada idioma es específico a una epistemología (Greenfield, 1997). Adicionalmente, la traducción implica que la versión de una prueba en el idioma original y la versión en el idioma objetivo han sido desarrolladas con procedimientos diferentes. La prueba en el idioma fuente ha sido desarrollada mediante un largo proceso iterativo que llega a incluir múltiples revisiones y el piloteo con muestras de estudiantes. En cambio, el proceso de traducción de pruebas se lleva a cabo en un tiempo relativamente corto, lo que limita la oportunidad de poner a prueba la traducción con estudiantes piloto y refinar la redacción de los ítems (Solano-Flores, Trumbull y Nelson-Barber, 2002).

#### Dimensiones del error de traducción

Debido a la inevitabilidad del error, nuestro modelo se ocupa especialmente de la clasificación detallada de errores de traducción. Para los propósitos de nuestra revisión de la traducción

de TIMSS-1995, misma que discutimos en la siguiente sección, desarrollamos un sistema de 10 dimensiones de errores de traducción (ver Tabla I). Las dimensiones obvias tienen que ver con exactitud de la traducción, la corrección gramatical (gramática, semántica) y las características editoriales y de producción de las pruebas traducidas (estilo, formato, convenciones). Otras dimensiones consideran el acuerdo de la traducción con el uso del idioma y con su usanza por la población destinataria, en sus contextos sociales e instruccionales (registro), el contenido (información, constructo) y la representación curricular (currículo). Una décima dimensión (origen) destaca el hecho de que las fallas del ítem en el lenguaje fuente pueden transferirse a la traducción.

Estas dimensiones de error no debieran ser interpretadas como exhaustivas o universales. Dependiendo de la naturaleza de las pruebas traducidas, cada dimensión puede tener una relevancia mayor o menor, o se le debe especificar con mayor o menor detalle de acuerdo con el contenido de la prueba, lo que ésta pretende medir y los recursos disponibles para el proyecto de traducción. Por ejemplo, mientras que en las pruebas TIMSS la traducción de términos técnicos tiene que examinarse con base en su alineación con el currículo implementado en grados escolares específicos, en las pruebas PISA es más pertinente examinarla con base en el uso del lenguaje a nivel nacional.

#### Relatividad de las dimensiones de error

Una traducción perfecta es prácticamente imposible de lograr, porque existe una tensión entre las dimensiones de error. Por ejemplo, usar un término clave en la traducción de un ítem más veces que en la versión original puede crear una diferencia en la cantidad de información crítica para entender un problema (dimensión de información); pero a la vez, esa inconsistencia puede ser necesaria para asegurar que exista una sintaxis natural según las convenciones en el uso del idioma de la traducción (dimensión de gramática).

#### Multidimensionalidad de los errores de traducción

Las propiedades lingüísticas de un ítem están interconectadas. Una consecuencia de esta interrelación es que una característica en la traducción de un ítem pueda pertenecer a más de una dimensión de error. Por ejemplo, la inserción inapropiada de una coma en una oración

puede violar algunas convenciones en la redacción del idioma objetivo (dimensión de gramática), pero también puede alterar el significado de una oración (dimensión semántica).

La traducción de un ítem no puede ser totalmente adecuada ni totalmente inadecuada. Es más apropiado pensar en la calidad de la traducción de un ítem en términos de su aceptabilidad. Debido a la multidimensionalidad de errores de traducción, prácticamente ningún ítem está libre de tener errores de traducción. Sin embargo, error no debe interpretarse como error fatal. Nuestro modelo postula que las personas tienen la capacidad cognoscitiva de manejar los errores de traducción de manera que, dentro de ciertos límites, éstos no afectan a su desempeño en las pruebas. A la vez, algunos errores mínimos que de manera separada podrían no tener importancia, pueden, en conjunto, afectar al desempeño de los examinados.

Naturaleza probabilística de la aceptabilidad de la traducción de un ítem

De acuerdo con nuestro marco conceptual, el análisis de los errores de traducción puede hacerse con un enfoque probabilístico. Los ítems lingüísticamente aceptables no están totalmente libres de error y no todos los ítems con errores son necesariamente cuestionables. La Figura 1 muestra la aceptabilidad (o cuestionabilidad) de la traducción de un ítem como un espacio probabilístico. Tal espacio probabilístico está delimitado por el número de errores de traducción (eje vertical) y por la severidad (eje horizontal) de esos errores (Solano-Flores, Contreras-Niño, y Backhoff-Escudero, 2005). La traducción de un ítem está en el área de aceptabilidad (área blanca), porque tiene pocos errores o porque tiene varios errores que son poco severos. A su vez, la traducción está en el área de cuestionabilidad (área sombreada), porque tiene pocos errores de traducción muy severos o porque tiene muchos errores que son poco severos.

### **2.30 La evaluación educativa**

La evaluación es hoy quizá uno de los temas con mayor protagonismo del ámbito educativo, y no porque se trate de un tema nuevo en absoluto, sino porque administradores, educadores, padres, alumnos y toda la sociedad en su conjunto, son más conscientes que nunca de la importancia y las repercusiones del hecho de evaluar o de ser evaluado. Existe quizá una mayor conciencia de la necesidad de alcanzar determinadas cotas de calidad educativa, de aprovechar

adecuadamente los recursos, el tiempo y los esfuerzos y, por otra parte, el nivel de competencia entre los individuos y las instituciones también es mayor.

Quizá uno de los factores más importantes que explican que la evaluación ocupe actualmente en educación un lugar tan destacado, es la comprensión por parte de los profesionales de la educación de que lo que en realidad prescribe y decide de facto el "qué, cómo, por qué y cuándo enseñar" es la evaluación. Es decir, las decisiones que se hayan tomado sobre "qué, cómo, por qué y cuándo evaluar". En general, uno de los objetivos prioritarios de los alumnos es satisfacer las exigencias de los "exámenes". En palabras de A. de la Orden (1989): "la evaluación, al prescribir realmente los objetivos de la educación, determina, en gran medida... lo que los alumnos aprenden y cómo lo aprenden, lo que los profesores enseñan y cómo lo enseñan, los contenidos y los métodos; en otras palabras, el producto y el proceso de la educación... querámoslo o no, de forma consciente o inconsciente, la actividad educativa de alumnos y profesores está en algún grado canalizada por la evaluación".

Todos estos factores han llevado a una "cultura de la evaluación" que no se limita a la escuela sino que se extiende al resto de las actividades sociales.

Concretamente, en nuestro país, la ampliación del ámbito de la evaluación desde los resultados y procesos del aprendizaje de los alumnos hasta el propio currículo (en sus distintos niveles de concreción), la práctica docente, los centros, el sistema educativo en su conjunto, etc, ha dibujado en los últimos años un nuevo escenario para las prácticas evaluativas, que se han desarrollado a todos los niveles de manera muy importante.

### **2.31 Naturaleza de la evaluación**

La evaluación, como concepto integrante de las políticas públicas, ha estado presente en el contexto educativo mexicano, en los últimos veinte años. El término ha estado vinculado con otro concepto: el de la calidad de la educación. A la evaluación se la concibe como una actividad indispensable y previa a toda acción conducente a elevar el nivel de la calidad de la educación. La evaluación así formulada constituye un momento de la planeación, entendida ésta como una acción racional dotada de propósito. Se identifica a la evaluación con la última etapa del proceso

natural del conocimiento que concluye con la emisión de juicios informados, proceso que antecede a las decisiones y a la acción humana.

En su forma más simple, la evaluación conduce a un juicio sobre el valor de algo y se expresa mediante la opinión de que ese algo es significativo. Se llega a este juicio calificando qué tan bien un objeto reúne un conjunto de estándares o criterios. Así, la evaluación es esencialmente comparativa. Supone la adopción de un conjunto de estándares y la especificación del grupo contra el cual el objeto es comparado. El objeto puede ser calificado como “bueno” o “malo”, cuando la referencia es la totalidad de los objetos o el objeto promedio del grupo; o bien como “mejor” o “peor”, si es comparado con un subconjunto particular del grupo. La evaluación, en esencia, supone adoptar un conjunto de estándares, definirlos, especificar el grupo de comparación y deducir el grado en el cual el objeto alcanza los estándares. Una vez realizado lo anterior, el evaluador está en posibilidad de hacer, en un segundo momento, un juicio sobre el valor del objeto evaluado.

Es preciso notar que no todas las evaluaciones son racionales y explícitas; los estudios de caso revelan los estándares, comparaciones y juicios de manera implícita e intuitiva, y algunos buscan entender más que juzgar. El juicio de valor puede ser expresado mediante palabras ordinarias de tipo descriptivo y no necesariamente mediante calificativos de “bueno” o “malo”. Los ejemplos y anécdotas en una descripción pueden servir para este propósito, ya que el juicio de valor reside en la forma en que las palabras son usadas.

El proceso de evaluación real es complejo y no se ajusta a una mera deducción a partir de ciertos estándares; más bien se trata de un proceso dialéctico entre principios abstractos y ejemplos concretos, en donde quien evalúa verifica sus juicios contra aquellas situaciones particulares en las que se siente confiado de conocer sus resultados. Aún más, los estándares pueden estar en conflicto unos con otros, y será necesario balancearlos así como establecer precedencias relativas entre ellos, algunas veces de manera intuitiva. Cuando el evaluador es distinto al que toma las decisiones, hay otro factor que interviene: la credibilidad. La preocupación del tomador de decisiones no es sólo atender al grupo de comparación y a los estándares, y cómo éstos son pesados relativamente entre sí, sino también si los hechos son reales y si están presentados por el evaluador en forma correcta.

Normalmente, la evaluación educativa supone la existencia de acuerdos comunes: juicios previos consensuados, sobre lo que es bueno o conveniente para todos como propósito de las acciones educativas. Este es el espacio que permite el disenso y la controversia sobre el sentido, el propósito y el uso de la evaluación en la educación. Es, asimismo, el que permite adjudicarle un carácter ideológico. La evaluación no es meramente un proceso técnico, neutral e inocuo; por el contrario, como toda acción social posee una dimensión política y, en su carácter instrumental, obedece a una lógica de control. Estos aspectos constitutivos de la evaluación están presentes, lo mismo en una actividad tan antigua y concomitante a los procesos de educación como la evaluación en el aula, como en las recientes políticas gubernamentales identificadas con procesos de metagestión, los cuales condicionan y vinculan recursos económicos a los resultados de la evaluación.

La evaluación asumida como un campo particular de la investigación educativa es reciente. Se ha cuestionado incluso si los proyectos de evaluación pueden considerarse investigación educativa. Se arguye que el mismo carácter instrumental de la evaluación contraviene el sentido desinteresado de la obtención de nuevo conocimiento que subyace a la lógica de la ciencia. Lo cierto es que, en un sentido amplio, el propósito de los proyectos de evaluación es la obtención de conocimiento, si bien se busca que éste sea útil para la toma de decisiones. La mayor parte de los estudios de evaluación requiere, asimismo, diseños metodológicos complejos y debe satisfacer normas que justifican la validez y confiabilidad del conocimiento que obtiene. En ello, las evaluaciones no se apartan de la parsimonia inherente a las investigaciones científicas.

Por otra parte, el fenómeno educativo, como el hombre mismo, es complejo y su conocimiento es siempre parcial e idiosincrático, pues está constreñido a la condición temporal de sus actores y posee una dimensión inevitablemente subjetiva. En este sentido, cualquier pretensión “científica” de las investigaciones educativas debe ser matizada, dada el carácter peculiar de su objeto de estudio. En lo que respecta a los estudios de evaluación, en los últimos años, se ha generalizado su aceptación como una rama de la investigación aplicada en educación.

### **2.32 Concepciones evaluativas**

El concepto de evaluación ha evolucionado durante los años, usualmente definiciones enfatizan las funciones tradicionales de medición que se le asigna a la evaluación dentro

del proceso educativo. Algunos enfoques están centrados en el control y la calificación, mientras que otras, asumen una perspectiva orientadora, formativa y en búsqueda de la mejora, apoyo y orientación de los procesos educativos.

El actuar profesional docente corresponde a todos los aspectos que deben ser cubiertos en la labor pedagógica, la cual se ha influenciado de los conocimientos de formación inicial y continua, pero además de la experiencia atesorada durante sus años de ejercicio. Este bagaje va determinando a lo largo de la carrera docente la forma de llevar a cabo, entre muchas otras cosas, el conocimiento del currículum, la planificación de la enseñanza, la reflexión y evaluación sobre su propia práctica, los efectos que tiene su forma de trabajar sobre el grupo de estudiantes.

Las concepciones evaluativas permean las acciones de los docentes en los procesos de evaluación, sin embargo, al no existir claridad en dichas concepciones epistemológicas se genera mayor resistencia a los cambios, en especial en la evaluación formativa. La evaluación educativa se ha transformado a lo largo de los años en cuanto a sus funciones y concepciones, originando diversos modelos vinculados a los movimientos sociales de la época.

Las adaptaciones que se pueden realizar en el proceso evaluativo buscan responder a las necesidades del estudiante en relación a qué, cómo y cuándo evaluar sus aprendizajes, encontrándose:

- Adaptaciones no significativas, las cuales apuntan a la modificar la selección o técnicas e instrumentos de evaluación.
- Adaptaciones significativas, las cuales apuntan a introducir criterios de evaluación específicos, así como modificar los criterios de promoción del estudiante de acuerdo a sus características.

Estas adaptaciones corresponden a recursos o estrategias con que cuenta los(as) docentes para llevar a cabo un desempeño profesional en un grupo que integra estudiantes con necesidades educativas especiales, lo cual exige al docente acercarse a los estudiantes para tener presente los antecedentes y conocimientos sobre las competencias que posee, así como los apoyos que requiere para avanzar en su proceso de aprendizaje, pudiendo implementar actividades de aprendizaje que promuevan la experimentación y el descubrimiento de nuevos esquemas cognitivos. El problema no se encuentra en las



características del estudiante sino que en las oportunidades y medios que favorecen su desarrollo

### **2.33 Funciones de la evaluación**

Tradicionalmente, la evaluación se ha venido aplicando casi con exclusividad al rendimiento de los alumnos, a los contenidos referidos a conceptos, hechos, principios, etc., adquiridos por ellos en los procesos de enseñanza. A partir de los años sesenta, la evaluación se ha extendido a otros ámbitos educativos: actitudes, destrezas, programas educativos, materiales curriculares didácticos, la práctica docente, los centros escolares, el sistema educativo en su conjunto y la propia evaluación.

Esta extensión de la evaluación a otros ámbitos tuvo lugar en los Estados Unidos a finales de los años 50 debido a circunstancias tales como: la crítica a la eficacia de las escuelas públicas, la gran inversión dedicada a la educación que exigía una rendición de cuentas (Accountability), etc. Por tanto, el campo de aplicación de la evaluación se extiende a alumnos, profesores, directivos, instituciones, la administración, etc. Y va a ser, precisamente, a raíz de la extensión del ámbito evaluador cuando van a surgir una serie de modelos de evaluación de gran relevancia.

### **2.34 Tipos de juicios evaluativos**

La Evaluación Iluminativa: es un modelo cualitativo que se fundamenta en dos ejes conceptuales: sistema de instrucción y el ambiente de aprendizaje. El sistema de instrucción está constituido por un conjunto de elementos que toma en cuenta y relaciona el entorno socio-psicológico y material, el proceso de la enseñanza por parte del docente y el aprendizaje por parte de los alumnos. El ambiente de aprendizaje está representado por un conjunto de relaciones culturales, sociales, institucionales y psicológicas, las cuales inciden en las normas, opciones y estilos de trabajo que tienen lugar en aula.

La Evaluación Respondiente: el propósito principal del modelo es responder a los problemas que se confrontan en el aula, descubriendo y ofreciendo un retrato completo y holístico de los procesos de enseñanza y de aprendizaje. Toma en cuenta las experiencias de cada uno de los

participantes, con el fin de responder a las necesidades de información y el nivel de conocimientos de quienes participan en el proceso.

La Evaluación Democrática: postula el derecho del alumno y del docente de expresar sus puntos de vistas y opciones en el aula, la escuela y la comunidad. Enfatiza el rol del docente evaluador como orientador que se ocupa de analizar y valorar el proceso del estudiante a través del uso de la formas de participación de la evaluación para transformar y modificar la práctica educativa.

La Evaluación Negociada: proponen el consenso de los distintos agentes que actúan en el proceso de evaluación, a través de ciclos dinámicos de conformación de ideas para tomar decisiones.

### **2.35 Objetivos, indicadores y estándares**

De acuerdo con Scheerens, Glass y Thomas (2005), los indicadores educativos son estadísticos que permiten realizar juicios de valor sobre la pertinencia de los aspectos clave del funcionamiento de los sistemas educativos; constituyen características mensurables de éstos y aspiran a medir sus aspectos fundamentales. Proporcionan un panorama de las condiciones actuales del sistema educativo, sin describirlo a fondo y se espera que a través de ellos sea posible establecer inferencias acerca de la calidad de la enseñanza. Debido a lo anterior, los indicadores educativos tienen, en general, como punto de referencia un estándar contra el cual pueden efectuarse los juicios de valor correspondientes.

Los indicadores educativos deben ser susceptibles de comparación a través del tiempo, esto es, deben dar cuenta del progreso y los cambios en el desempeño de cierta variable; en otras palabras, el indicador no sólo cumple una función informativa, sino también evaluativa, pues se espera que, mediante ellos, sea posible determinar si existen mejoras o deterioros en algunas variables del sistema educativo (Morduchowicz, 2006). A este respecto, Kanaev y Tuijnman (2001), citados en Morduchowicz (2006), señalan que, además de su función informativa, los indicadores permiten construir nuevos enfoques y expectativas. Cabe destacar que los indicadores no sólo proporcionan información considerando el contexto del sistema, sino que facilitan el análisis de tendencias y la proyección de situaciones futuras del mismo sistema.

Según Morduchowicz (2006), desde la perspectiva de las políticas públicas, la selección y el uso de indicadores no es una tarea neutra; tiene dos dimensiones que permiten su análisis: la dimensión técnica y la dimensión política. Los indicadores pueden, entonces, referirse no sólo a las políticas, sino a las características generales del sistema educativo.

Los indicadores, en cuanto que constituyen propuestas de interpretación de la realidad, no pueden ser entendidos como herramientas capaces de proporcionar una visión acabada de la realidad tal cual es. En otras palabras, difícilmente permiten la comprensión de un fenómeno en toda su magnitud y complejidad, pues sólo aportan un marco de referencia cuantitativo que no incorpora los elementos cualitativos del fenómeno (Morduchowicz, 2006). Es comprensible, por tanto, que utilizar un solo indicador para obtener información acerca de un fenómeno sea inadecuado. Los indicadores no son cifras aisladas; se encuentran interrelacionados unos con otros, y para obtener una clara comprensión de la información que proporcionan deben agruparse y constituirse en lo que se conoce como sistema de indicadores.

Los sistemas de indicadores facilitan la descripción de situaciones que no pueden medirse de manera directa; por ejemplo, medir constructos tales como la calidad de la docencia o la de la educación puede resultar sumamente difícil, pues no existe un solo indicador que por sí mismo refleje este constructo. En cambio, si miden los mismos constructos considerando una serie de indicadores tales como formación académica, experiencia laboral, o resultados de aprendizaje y eficiencia terminal, sería posible configurar un panorama más claro en relación con estos constructos.

De acuerdo con Ogawa y Collom (1998), los sistemas de indicadores pueden adoptar dos características principales que hacen referencia al número de indicadores que conforman un sistema: indicadores parsimoniosos y extensos. Estos últimos incluyen un gran número de indicadores en su conjunto, y han sido criticados por resultar inmanejables y complejos; por su parte, los parsimoniosos contienen un reducido número de indicadores y las críticas que han recibido se relacionan, principalmente, con su poca capacidad para dar cuenta con eficacia de la complejidad de fenómenos que configuran al sistema educativo (Ogawa y Collom, 1998).

El sistema de indicadores educativos que publica la OCDE desde 1998 con el título de Education at a Glance, constituye un ejemplo de un sistema de monitoreo que pretende

establecer comparaciones entre diferentes países respecto a la situación del sistema educativo. Education at a Glance incluye las visiones más actuales sobre la construcción y el cálculo de indicadores en diversos dominios educativos. Su publicación ha generado una amplia red de especialistas y un conjunto relevante de conocimientos (CERI, 1994), que ha impactado el desarrollo y la evolución de los sistemas de evaluación y monitoreo en diversos países a escala mundial. En el contexto nacional se ha iniciado un proceso semejante a partir de la publicación anual del Instituto Nacional para la Evaluación de la Educación (INEE) denominada Panorama educativo: indicadores del Sistema Educativo Nacional, que se edita desde 2003 para presentar los resultados del diseño y la aplicación de un sistema de indicadores de contexto, insumo, proceso y producto.

### **2.36 La evaluación como proceso**

Cuando se habla del proceso continuo de la evaluación, nos estamos refiriendo a una “recolección permanente de información”, lo cual se refiere a darle seguimiento al desempeño durante todo el período escolar, pensando en impulsar la mejora continua, lo cual sólo es posible si hay reflexión permanente del proceso de enseñanza aprendizaje y seguimiento al desempeño.

La continuidad en el proceso permite igualmente recolectar información representativa de los evaluados, y no solamente observaciones aisladas y puntuales para emitir valoraciones ajustadas al desempeño real. Esas aproximaciones sucesivas ponen el carácter continuo del proceso y obligan a que dicho proceso de recuperación de evidencias entre a desempeñar papeles protagónicos dentro de la evaluación.

Hay que cuidar que este proceso de recolección de datos de manera continua no se limite solamente a la integración inerte de información. Es importante recoger evidencias de lo que los alumnos presentan un cambio de postura en el proceso de evaluación, porque si la finalidad es lograr una actuación competente, esta información nos permite planear estrategias dirigidas y luego búsquedas más autónomas, dándole seguimiento a la forma en que los alumnos van haciendo aportaciones (entregando evidencias).

## 2.37 Componentes del proceso evaluativo

Vimos que evaluar es juzgar, es decir, emitir un juicio valorativo. En evaluación de los aprendizajes, los juicios son descriptivos, pues, el docente emite una opinión crítica que expone el desempeño del estudiante, precisando la actividad realizada. En tano juicio, un juicio evaluativo o juicio de valor descriptivo tiene la forma lógica de una proposición; o sea, está configurado por un sujeto y un predicado, unidos por un verbo copulativo como los verbos ser y estar. Veamos estos elementos por separado.

**El Sujeto.** Es aquel (persona o cosa) de quien se afirma o se niega algo. En la evaluación dentro del ámbito educativo, el sujeto es el educando, aquel ser sobre el cual recae el proceso de enseñanza-aprendizaje y cuyos logros se pretende valorar. Cabe señalar que se valora y describe su comportamiento, desempeño, habilidades, actitudes; no la persona. Es incorrecto expresar: “Elena es mala estudiante”; mejor y correcto es: “Elena está en el proceso de alcanzar los objetivos previstos” (en forma positiva), sin incorporar preferencias o sentimientos.

**Cópula.** Es el verbo que une al sujeto con el predicado y expresa la afirmación o negación que se realiza sobre el sujeto; informa lo que hace o le sucede al sujeto. Generalmente está conformada por los verbos ser y estar, aunque puede ser otro verbo. No obstante, el verbo utilizado siempre remite a la actuación o realización específica que describa el juicio y que exprese las acciones que realiza el sujeto. En otras palabras, da a conocer la acción o movimiento, existencia, consecución, condición o estado del sujeto. Semánticamente expresa una predicación completa.

**Predicado.** En evaluación el predicado está conformado por todo aquello que se afirma o se niega del sujeto, es lo que se dice del sujeto. Con base en la información acopiada sobre la actuación del estudiante y de los criterios previstos, contiene y describe comportamientos, desempeño, objetivos educativos logrados, aprendizajes y conocimientos alcanzados, potencialidades desarrolladas; o sea, las competencias que el sujeto ha logrado adquirir y/o desarrollar mediante el proceso de aprendizaje. En el predicado se reflejan las fortalezas, debilidades, avances del estudiante durante el período que se evalúa

## 2.38 La metaevaluación

La metaevaluación es una composición conceptual. Proviene del prefijo griego Meta que significa “más allá”, y del concepto evaluar. Fue desarrollado por Michael Scriven, quien la definió como: “La evaluación de las evaluaciones (Scriven, M, 1991) Su objetivo es asegurar la calidad de los servicios evaluativos, evitar enfrentarse a las prácticas ilegales y a los servicios que no son de interés público, señalar el camino para el perfeccionamiento de la profesión y promover una mayor comprensión de la institución evaluada (Stufflebeam, D, et al, 1987). Busca verificar la información, los juicios de valor y promover cambios o la continuidad de un proyecto o programa.

Podemos ver que el fin, tanto de la evaluación como de la metaevaluación, son afines, entonces ¿cuál es la diferencia? ¿Por qué someter a juicio las evaluaciones?

¿Por qué evaluar las evaluaciones? “toda acción tiene una segunda y hasta una tercera intención”. Evaluamos para mejorar la equidad y calidad de los programas e instituciones, pero en ocasiones, detrás de las evaluaciones se esconden intereses económicos, políticos, sectoriales etc. En este sentido la metaevaluación sería como un desvelamiento de lo visible, una mirada más allá de las evidencias primeras, una evaluación de la evaluación.

Distintos autores se han percatado de estas dobles intenciones y se han atrevido a iniciar la tarea de descubrir e interpretar para comprender lo que está más allá de lo visible en las prácticas evaluativas, tratando de desentrañar el significado de algunas acciones que suceden en las evaluaciones. La metaevaluación trata de develar y des-ocultar el significado profundo de las mismas. Además, nos permite tener datos empíricos para juzgar los resultados generales de la evaluación de los programas. En este sentido también es un instrumento de validación de las prácticas científicas de los evaluadores y del modelo de los estándares desarrollados.

Gustavo Laies afirma que la evaluación de la evaluación debido a su grado de efectividad ha dejado de ser una herramienta de uso exclusivo de la educación y se ha convertido un procedimiento de las políticas públicas en su conjunto. Revela las desigualdades sociales, la competitividad de los países, la eficiencia del gasto, entre otras variables (Laies, G, 2003)

Por otra parte, la metaevaluación tiene un talante dialógica. En tanto que los resultados se vuelven una noticia y adquieren un fuerte reconocimiento social y pueden ser insumos para el debate del cual pueden salir ideas para futuros programas consensuales.

Por último creo que también la metaevaluación tiene un componente existencial que se puede afirmar como una búsqueda de sentido a través de las acciones transparentes. Es fundamental recuperar los modos de construir políticas de sentido con fundamentos que puedan sostener los acuerdos técnicos posteriores.

En síntesis, podemos decir que la función de la metaevaluación es analizar y evaluar los procesos de evaluación. Se evalúa para mejorar la equidad y calidad de nuestras instituciones y sistemas educativos, con evidencias de evidencias se hace más creíble juzgar al sistema o una acción. Evaluar constantemente las evaluaciones debería ser un deber ser de toda institución.

En educación es primordial para el desarrollo de todo sistema que aspira a la mejora continua. La educación que no se cuestiona, que no recoge cada cierto tiempo sus programas para re-evaluarlos y devolverlos renovados hacia la mente de sus estudiante no podrá ver más allá de lo visible. La metaevaluación es eso: evolución y cambio, es una meta-mirada capaz de ver lo que subyace más allá de lo aparente. Esa es la finalidad última de la metaevaluación, por eso hay que evaluar las evaluaciones, para mirar más allá de lo visible.

### **2.39 Algunos tópicos de interés acerca de la valuación**

En la reforma la evaluación se define como un elemento y proceso básico en la práctica educativa que orienta y permite efectuar juicios de valor necesarios para orientar y tomar decisiones en referencia al proceso de enseñanza-aprendizaje.

Como proceso la evaluación forma un continuo a lo largo de todo el proceso de enseñanza-aprendizaje del alumno. No se ha de limitar a situaciones aisladas donde se aplican unas pruebas y se obtienen unos resultados que son anotados. La evaluación se ha de realizar en todo momento, y por medio de la reflexión sobre lo que se hace.

La función orientadora de la evaluación señala los aspectos sobre los que hay que intervenir con el fin de mejorar el proceso de enseñanza-aprendizaje.

Para realizar la evaluación se recoge información de los cambios producidos en el desarrollo educativo del alumno, la adecuación de los programas y actividades puestas en práctica, la eficacia de la metodología y de los recursos empleados.

La toma de decisiones forma un elemento esencial del proceso evaluador, con el fin de mejorar el proceso de aprendizaje, cambiar el proyecto diseñado por el docente, incorporar los mecanismos de corrección precisos, programar el plan de refuerzo, intervenir para resolver conflictos actitudinales, orientar la acción tutorial, entre otros elementos.

Por ello la evaluación no aparece centrada exclusivamente en el alumno y se extiende a todo el entorno educativo que lo rodea.

#### **2.40 Ámbitos básicos de la evaluación educativa**

Tradicionalmente, la evaluación se ha venido aplicando casi con exclusividad al rendimiento de los alumnos, a los contenidos referidos a conceptos, hechos, principios, etc., adquiridos por ellos en los procesos de enseñanza. A partir de los años sesenta, la evaluación se ha extendido a otros ámbitos educativos: actitudes, destrezas, programas educativos, materiales curriculares didácticos, la práctica docente, los centros escolares, el sistema educativo en su conjunto y la propia evaluación.



## UNIDAD III

### LA EVALUACION DE LOS APRENDIZAJES DE LOS ALUMNOS

#### 3.17 Cambios fundamentales en la evaluación de los aprendizajes de los alumnos

Un primer motivo que explica el auge actual de la evaluación es el cambio registrado en los mecanismos de administración y control de los sistemas educativos, que ha marchado paralelo a las propias transformaciones experimentadas por el aparato escolar en las últimas décadas. Cualquier observador que analice la situación actual de los sistemas de educación y formación apreciará que éstos se encuentran sometidos a exigencias crecientes por parte de los ciudadanos, las sociedades y las administraciones. Entendidos como instrumentos al servicio del cambio social, del desarrollo económico y de la producción del conocimiento, los sistemas educativos son objeto de demandas múltiples y progresivamente más exigentes.

De ellos se espera que contribuyan a crear sociedades abiertas, activas y equitativas; economías dinámicas y competitivas; sistemas políticos democráticos y pluralistas; personas, en fin, equilibradas, tolerantes y socialmente integradas. En consecuencia, las demandas que reciben dichos sistemas con cada vez mayores, produciendo como efecto el replanteamiento de muchos de sus principios tradicionales.

En una época de transformaciones tan aceleradas como la actual y de demandas tan exigentes sobre el aparato escolar como las expuestas, la rigidez de un sistema educativo es considerada un elemento de retraso, más que de progreso. Ello conduce a enfatizar muy especialmente características tales como su flexibilidad o su capacidad para adaptarse y dar respuesta a las necesidades del desarrollo de las naciones y de las sociedades. El corolario, que parece inevitable, es la concesión de un cierto grado de autonomía a los centros educativos para que puedan regular su propia vida en un marco de libertad, adaptando así su oferta a las necesidades de su entorno. Obviamente, los límites dicha autonomía y los grados de libertad concedidos a las escuelas varían notablemente de unos lugares a otros: los límites del fenómeno comúnmente conocido como descentralización son muy variables, según han puesto de manifiesto diversos estudios (PUELLES, 1992; MUR, 1993). Pero, más allá de esa apreciable

diversidad, cabe reconocer que las características generales del proceso apuntado son cada vez más evidentes a escala tanto nacional como internacional. No parece osado aventurar que los cambios en la administración y el control de los sistemas educativos derivados de dicho proceso no serían meramente coyunturales, sin que tendrán importantes consecuencias para el futuro.

La principal transformación registrada se refiere tanto al modo de ejercer el control como al objeto del mismo. Así, mientras que los mecanismos de administración y control en sistemas educativos más estables que los actuales (y especialmente en aquellos con estructuras centralizadas) han sido de tipo jerárquico y centrados en los procesos que tienen lugar en las escuelas, los correspondientes a sistemas flexibles y dotados de una cierta autonomía interna son generalmente más participativos y centrados en los resultados. Este cambio es fundamental para entender cómo se organiza y regula un número cada vez mayor de sistemas de educación y formación, hecho que tiene importantes implicaciones para su funcionamiento.

En efecto, el control y la administración tradicionales estaban basados en dos procesos centrales: en primer lugar, la elaboración de normativa, de diverso rango, alcance y cobertura; en segundo lugar, el control del cumplimiento de las normas generalmente a través de servicios jerárquicos de supervisión e inspección. En sus casos más extremos, la actuación de dichos servicios llegaba al punto de pretender conocer con exactitud las actividades desarrolladas diariamente de manuales, la fijación del horario y de la distribución de las materias, el establecimiento de programas detallados y la realización de visitas regulares de inspección.

En la actualidad, esos sistemas de control resultan cada vez más chocantes y contradictorios con el modelo de organización escolar que antes se apuntaba. No resulta fácil articular un modelo escolar descentralizado, en el que el control de su actuación tiende a revertir sobre la comunidad a la que sirve, con unos mecanismos jerárquicos de administración y gestión. En este nuevo contexto, la participación democrática en el control del sistema educativo se convierte en una exigencia insoslayable. Ya no es la Administración la única competente para controlar la vida escolar, sino que los ciudadanos y las sociedades deben tomar parte en dicha tarea. A ello se hacía mención al hablar de un control y una gestión participativos.

Por otra parte, en este nuevo modelo de organización escolar, la inspección y supervisión de los procesos dejan paulatinamente paso a la valoración de los resultados, lo que implica la consiguiente puesta en marcha de procesos de evaluación. Las escuelas pueden y deben hacer uso del creciente margen de autonomía de que disponen, organizando los procesos de enseñanza y aprendizaje en ese nuevo marco. Pero, a cambio, deben rendir cuentas de sus resultados, a través de diversos mecanismos de evaluación. Es éste un estilo de funcionamiento frecuente en las organizaciones productivas, que ha ido arraigado también entre las administraciones públicas. Parece que la educación no constituye una excepción a lo que va siendo una regla cada vez más extendida en la gestión de los servicios públicos.

No obstante, la valorización de los resultados de la educación es una tarea compleja, que entraña algunos riesgos. Deberían señalarse aquí al menos, de ellos. El primero que es el de reducir dicha valoración a los elementos más asequibles, al margen de su relevancia. Cualquier sistema de evaluación a los elementos más asequibles, al margen de su globalidad, no reduciendo la tarea educativa a la meramente instructiva, ni está a la cognoscitiva. El segundo riesgo es el de utilizar dicha evaluación para realizar comparaciones y clasificaciones injustas a los centros. Los sistemas utilizados deben ser muy sensibles respecto de sus posibles usos y efectos perversos. Pero la constatación de tales riesgos no invalida la dirección emprendida por un número creciente de sistemas educativos hacia la evaluación de los resultados de la educación. En resumen, podría decirse que esta parece ser una tendencia con visos de consolidación en el futuro próximo.

### **3.18 Modelos de evaluación de los aprendizajes**

Este modelo contiene dos grandes apartados; por un lado el que se relaciona de la evaluación, que como su nombre lo indica es la referencia que debemos tener presente al planear y desarrollar la evaluación. En este modelo, el referente lo constituyen los programas por competencias.

El segundo gran apartado es la evaluación, que como observamos, tiene los siguientes componentes: desempeño o actuación (ejecución) del estudiante; los criterios, que son los que califican a los desempeños en cuanto a sus características de calidad y de enfoque pedagógico;

los productos del aprendizaje, que son la transformación y reconstrucción de los conocimientos durante una experiencia de E-A-E y la capacidad de transferirlo en productos; las evidencias que son la expresión objetiva y concreta de los productos del aprendizaje y, finalmente los juicios de valor que son la reflexión y valoración que se hace para determinar si los desempeños y competencias han sido logradas.

Asimismo, al ir desarrollando el modelo fuimos conceptualizando cada uno de los componentes del mismo y le fuimos dando forma para hacerlo aplicativo en los diversos ambientes de aprendizaje. Así, en cuanto a los referentes de la evaluación, que corresponden a los programas de las unidades de aprendizaje o cursos que van a ser evaluados, es conveniente que se elaboren y se presenten por CPI, y que el desarrollo para la adquisición de las competencias se programe por procesos, en lugar de acciones o estrategias de enseñanza-aprendizaje.

### **3.19 Procedimientos tradicionales de evaluación de los aprendizajes**

La evaluación del proceso de enseñanza permite también detectar necesidades de recursos humanos y materiales, de formación, infraestructura, etc. y racionalizar tanto el uso interno de estos recursos como las demandas dirigidas a la Administración para que los facilite en función de las necesidades.

Es importante resaltar que la evaluación de la propia práctica docente, bien sea de forma individual o del conjunto del equipo, se muestra como una de las estrategias de formación más potentes para mejorar la calidad del proceso de enseñanza-aprendizaje. Por otra parte, la evaluación del equipo docente en su conjunto permite detectar factores relacionados con el funcionamiento de la coordinación, las relaciones personales, el ambiente de trabajo, aspectos organizativos, entre otros que son elementos muy significativos en el funcionamiento de los centros.

Esta evaluación de la intervención educativa debe hacerse en dos niveles distintos o ámbitos:

- a) El contexto del aula
- b) el conjunto del centro

En el primer caso el responsable es cada maestro, mientras que en el ámbito del centro lo es el conjunto del profesorado.

Veamos uno a uno:

#### 1) La Práctica docente en el contexto del Aula:

- \* El diseño y desarrollo de la Unidades didácticas y la adecuación de las adaptaciones realizadas para grupos de alumnos concretos.
- \* El ambiente del aula y todo aquello que favorezca el proceso de enseñanza y aprendizaje: organización, espacios y agrupamientos.
- \* La actuación personal de atención a los alumnos.
- \* La coordinación con otros maestros que intervienen en el mismo grupo de alumnos.
- \* La comunicación con los padres.

#### 2) La Práctica docente en el contexto del Centro:

##### A) En el Ciclo:

- \* Eficacia del sistema e instrumentos de coordinación establecidos.
- \* Adecuación de los criterios de evaluación y promoción
- \* Coherencia interna del ciclo B) En la Etapa:
  - \* Los elementos de la Programación y su coherencia.

La coordinación interciclos y coherencia vertical entre cada área.

- \* Las medidas de atención a la diversidad adoptadas
- \* La relaciones entre las áreas
- \* El tratamiento de los temas transversales

##### C) En el contexto del Centro:

- \* La coherencia del Proyecto Educativo.

- \* La funcionalidad de decisiones sobre asignación de recursos humanos, materiales, espacios y tiempos.
- \* El funcionamiento de los órganos unipersonales y colegiados
- \* Las relaciones con familias y entorno
- \* Los servicios de apoyo (EOEPs, SITE, CPR)
- \* La utilización de los recursos de la comunidad.

### **3.20 Procedimientos alternativos de evaluación de los aprendizajes**

Con la aparición de nuevos desarrollos en el campo de la evaluación han traído a la educación un nuevo campo: lo que se conoce como evaluación alternativa y se refiere a los nuevos procedimientos y técnicas que pueden ser usados dentro del contexto de la enseñanza e incorporados a las actividades diarias el aula

Existen diferencias entre la evaluación tradicional y la evaluación alternativa, esta última permite:

Centrarse en documentar el crecimiento del individuo en un determinado tiempo, en lugar de comparar a los estudiantes entre ellos mismos.

Reforzar las cualidades y destrezas antes que las debilidades de los estudiantes.

Tener en cuenta los estilos de aprendizaje, las capacidades lingüísticas, las experiencias culturales y educativas y los grados de estudio.

Los eruditos en el tema nos indican que los exámenes tradicionales de respuesta fija no den una visión clara y veraz sobre lo que los estudiantes pueden traer con sus conocimientos, solamente permiten traer a la memoria, observar la comprensión o interpretación del conocimiento, pero no demuestran la habilidad del uso del conocimiento. A la par, se argumenta que los exámenes estandarizados de respuesta fija ignoran la importancia del conocimiento holístico y la integración del conocimiento y, no permiten evaluar la competencia del estudiante en objetivos educacionales de alto nivel de pensamiento o de lo que espera la sociedad. Además, con frecuencia el resultado de las evaluaciones se emplea solamente para

otorgar una nota a los examinandos y no reingresa en las estrategias de enseñanza y de aprendizaje para mejorar los esfuerzos.

En este sentido debemos recordar que la evaluación alternativa incorpora una variedad de técnicas de evaluación, entendiendo estas como “cualquier instrumento, situación, recurso o procedimiento que se utilice para obtener información sobre la marcha del proceso” (Silva, 2018, p.246); dichas técnicas se pueden adaptar a diferentes situaciones. Existen 2 clases de alternativas, las técnicas para la evaluación del desempeño (mapas mentales, solución de problemas, método de casos, proyectos, diarios, debate, ensayo, técnica de la pregunta y portafolios) y las técnicas de observación (entrevista, lista de cotejo, escalas, rúbricas, ) estas últimas son un complemento para las primeras.

### **3.21 A modo de conclusión**

La evaluación del proceso de enseñanza-aprendizaje en Ciencias Sociales se encuentra estrechamente relacionada a la perspectiva teórica de la cual parte el docente y, por ende, a las finalidades que le asigna a la enseñanza de las demás ciencias. Muchas decisiones sobre la enseñanza y, en particular sobre la evaluación, omiten la teoría y se basan en rutinas o en opiniones que no se explicitan claramente o ignoran el marco teórico desde el cual parten, por eso, es común observar contradicciones y también frecuentes errores. Por eso, planteo analizar el sustento ideológico y científico desde el que partimos, para poder tomar decisiones consientes acerca de porqué enseñar Ciencias Sociales y cómo orientar ese proceso de enseñanza y cuándo y de qué manera evaluarlo. Se suele perseguir una interpretación cuantitativa del proceso educativo a partir del manejo y ordenamiento de datos, el establecimiento de causas y consecuencias de un determinado acontecimiento. De tal forma que estamos en la presencia en presencia de un paradigma de corte positivista, que concibe a la evaluación en términos de éxito o fracaso, en la medida en que los alumnos logran o no logran alcanzar un saber válido, confiable y aplicable. Cuando se evalúa para medir sólo se pretende controlar el desempeño o actuación de los estudiantes.

Reconceptualizando la evaluación como se ha propuesto permite al docente el seguimiento del desarrollo de las actividades por parte de los estudiantes, responder a sus consultas, detectar

las dificultades que se plantean, realizar ajustes a las actividades presentadas, analizar las formas en que los alumnos resuelven las situaciones planteadas y cómo construyen y aplican conceptos estructurantes y específicos del área y también cómo presentan sus producciones –orales y escritas- utilizando un vocabulario apropiado, comunicar a los estudiantes las metas que alcanzaron en relación con la situación de partida, verificar si se han conseguido las intenciones educativas.

Los estudiantes poseen espacio para la creatividad, el despertar la necesidad de la búsqueda de información a partir de diversas fuentes, la lectura, análisis y confrontación de las mismas, organizar el trabajo y las responsabilidades de cada uno, aplicar lo aprendido precedentemente y también la nueva información, ampliar el sentido de ciertos conceptos y aproximarse a otros conceptos, desarrollar la libertad de elección atendiendo al razonamiento como fundamento de la misma, analizar los errores para poder superarlos, desarrollar mecanismos de autoaprendizaje .

La evaluación formativa se puede llevar adelante a partir de la observación, donde el docente se incorpore el trabajo individual del estudiante, así como también su desempeño grupal (se integra, colabora, organiza, se interesa, emprende la búsqueda de materiales que profundicen los contenidos abordados) y la presentación oral y escrita de los trabajos que se requieran. Una alternativa para la autoevaluación del docente y los alumnos registrar todo lo que vaya sucediendo en el proceso de enseñanza-aprendizaje en un cuaderno que puede titularse diario del docente y diario del alumno, respectivamente. La escritura, lectura y reflexión de lo que allí se vuelque, permitirá realizar los cambios necesarios para mejorar la acción educativa.

### **3.22 Modelos de evaluación alternativa: La evaluación en el contexto del ABP y del portafolios**

Enseñanza y evaluación son las dos caras de una misma moneda. Entre ambas debe haber una correlación absoluta: nuestra manera de enseñar debe determinar nuestra manera de evaluar porque, de manera inevitable, nuestra manera de evaluar condiciona la manera de aprender de nuestros estudiantes.



Por ello, si nuestro estilo de enseñanza está cerca de la "instrucción directa" que describiremos en el primer módulo y nuestro estilo de evaluación consiste en un examen escrito en el cual solicitamos de nuestros estudiantes que reproduzcan el contenido que nosotros expusimos en la fase de presentación, el mensaje en relación con el estilo de aprendizaje de nuestros estudiantes es bien claro: es necesario memorizar los contenidos y ninguna otra competencia (en inglés podríamos hablar de 21st Century Skills) es necesaria.

Si, por el contrario, tenemos intención de valorar cómo se aprende y qué se aprende durante el desarrollo de nuestro proyecto y al final del mismo, así como la calidad del producto final y otros aspectos que encontremos relevantes (relaciones entre los estudiantes, relaciones dentro del grupo, cambio actitudinal, etc.), entonces tenemos necesidad de buscar mecanismos de evaluación alternativos más allá de las herramientas convencionales: estamos hablando de una evaluación alternativa para una enseñanza alternativa.

#### Definición de conceptos fundamentales

La evaluación es un tema central de la profesión docente. En primer lugar, la evaluación es una práctica de gran relevancia en la vida de nuestros estudiantes y en su aprendizaje. Por un lado, las decisiones que se adopten a partir de los resultados de la evaluación pueden determinar aspectos centrales de sus vidas (progresar o no progresar en los estudios, elegir unas opciones u otras, tener acceso a becas o puestos de trabajo determinados, etc.). Por otro lado, cómo evaluamos y qué resultados ofrecemos a los estudiantes a partir de la evaluación determina, en gran medida, cómo aprenden nuestros estudiantes, como ya hemos comentado.

La educación superior ha estado trabajando en los últimos años con la idea de que los estudiantes tienen que graduarse con una mayor capacidad para integrarse en su entorno disciplinar y construir conocimiento acerca del propio desempeño (Jones, 2010; Taylor, 2011). Lo que caracteriza a cualquier profesional es su habilidad para resolver problemas nuevos, analizar, enjuiciar y tomar decisiones y participar en la investigación respecto de su propia práctica con el objetivo de mejorarla (Jones, 2010; Manriquez Pantoja, 2012) . En este sentido, si la educación superior se dirige a la formación de profesionales competentes debe orientar sus estrategias hacia el impulso de esta reflexión concerniente con la acción, y ello implica

replantear el aprendizaje y la evaluación e interpretar la segunda como una actividad significativa, justa y transparente (Admiraal, Hoeksma, van de Kamp & van Duin, 2011).

Esta visión del aprendizaje que afecta a todas las etapas del sistema educativo ha empezado a calar hondo en la institución universitaria europea a raíz del denominado proceso de Bolonia. La universidad ha tenido que afrontar dos grandes retos. Primero, la dificultad de identificar y definir los resultados de aprendizaje más relevantes en un mundo en constante transformación. Segundo, cómo promover una participación más profunda y responsable del estudiante en la construcción de su proceso de aprendizaje a lo largo de la vida y en la aplicación del conocimiento en los diferentes entornos del mundo real (Taylor, 2011).

Lograr el justo término entre tutelar el aprendizaje, centrarlo en hechos que acontecen dentro y fuera del aula, despertar el interés de los estudiantes y respetar y potenciar la singularidad de cada uno de ellos es posiblemente el eje alrededor del cual se ha concebido el nuevo perfil de estudiante y el nuevo concepto de evaluación formativa (Knight, Hakel & Gromko, 2008; Peacock, Gordon, Murray, Morss & Dunlop, 2010). La participación del estudiante en el proceso de evaluación, el uso de múltiples medidas capaces de reflejar y estimular las también múltiples capacidades cognitivas y el impulso de una evaluación auténtica integrada en el propio proceso de aprendizaje son algunas de las líneas de fuerza que concretan el nuevo paradigma de evaluación que está arraigando en la educación superior (Baeten, Dochy & Struyven, 2008).

Como señalaron entre otros Withaus (2002), Orzoff, Peinovich & Riedel (2008), o Graham & Sims-Gunzenhauser (2009), la evaluación no solo tiene que intentar reflejar la capacidad del estudiante para producir artefactos propios de un ámbito disciplinar determinado, sino que debe plantearse como un auténtico reto para que aquel pueda tomar conciencia del aprendizaje construido. Valorar hasta qué punto un proceso evaluador que exige negociación, realimentación y consideración del contexto concreto en el que se produce no reclama asimismo una participación destacada del estudiante es algo delicado. Y no únicamente en el sentido de que ello significa transformar el rol del docente, del discente y de la institución educativa en su conjunto, sino también en el sentido de que esto sucede en la última etapa del proceso formativo: precisamente en el momento en el que se construye al profesional a punto de ejercer.

Ante tales retos, es urgente un enfoque curricular integral en el que el aprendizaje autónomo y la capacidad de los estudiantes para resolver problemas se sitúen al centro de la acción educativa. El Aprendizaje Basado en Problemas (ABP en adelante) constituye un ejemplo de este nuevo enfoque; de hecho, surgió en la década de los años setenta del siglo XX para afrontar los nuevos desafíos que planteaba la medicina (Font Ribas, 2004; Bermejo y Pedraza, 2008). Su rápida extensión por los Estados Unidos y poco después por Europa logró dejar atrás las reticencias iniciales (Branda, 2009) y postularse como una de las innovaciones más radicales en el entorno académico de las ciencias de la salud (Boud & Feletti, 1998). El principio del ABP sitúa al estudiante en el centro del proceso de aprendizaje, desplaza la figura del profesor hasta convertirle en un tutor facilitador y plantea una situación problemática que el primero debe resolver con el apoyo del segundo (van Berkel & Schmidt, 2000; Dolmans & Ginns, 2005; Woltering, Herrler, Spitzer & Spreckelsen, 2009).

### **3.23 Modelos alternativos de evaluación de los aprendizajes, principios caracterizadores**

Es aquella que se integra al proceso de enseñanza-aprendizaje, de manera que deja de ser un momento final, separado e independiente. Cuando la evaluación de los aprendizajes tiene un propósito formativo bien pensado y se planifica con ese fin, podría convertirse en uno de los elementos que dan dirección y sentido al diseño de materiales, a la programación de cursos, a la acción de facilitación y apoyo al aprendizaje y, en última instancia, a la formación de un estudiantado autónomo capaz de seguir aprendiendo solo. Esto sucedería porque dicha evaluación formativa estaría supeditada al proceso de aprendizaje, como un mecanismo que no se usa para medir, sino para construir aprendizaje significativo.

La evaluación formativa hace referencia a todo un proceso de evaluación cuya finalidad principal es mejorar los procesos de enseñanza-aprendizaje. Es todo proceso de aprendizaje que sirve para que los estudiantes aprendan más (y/o corrijan sus errores) y para que el profesorado aprenda a trabajar mejor (a perfeccionar su práctica docente). Siguiendo a López (2009), en la literatura inglesa se suele utilizar el término “formative Assessment” el término assesement proviene de una raíz latina cuyo significado es “sentarse al lado de “, sentido de dar ayuda o cooperar con, en vez del significado de inspección y control, que es la forma habitual de

entender y practicar la evaluación. Este tipo de evaluación formativa cambia los roles de quienes actúan en el proceso, y se vuelven más dinámicos y cohesionados, pues la idea es que un modelo de evaluación formativa promueva inteligentemente y con propósitos claros la participación de todas las personas involucradas. Algunas estrategias para incluir la evaluación formativa dentro del proceso de enseñanza-aprendizaje podrían orientarse de la siguiente manera:

Desde el docente La guía y la retroalimentación del tutor(a) en cada una de las actividades de aprendizaje en las que participa el estudiantado (ya sea en un foro, en una tarea o en un trabajo colaborativo), de manera que se convierten en objeto de estudio y mejoramiento permanente.

Desde los mismos estudiantes La promoción de la autorregulación del aprendizaje que debe estar pensada mediante una estrategia cuidadosamente diseñada desde el inicio del curso, de tal forma que el estudiantado posea de antemano los criterios que le ayudarán a autorregular su propio proceso de aprendizaje. Esto ha de ser una preocupación transversal.

Desde los(as) compañeros(as) La coevaluación entre pares desde el trabajo colaborativo y la construcción conjunta de conocimiento, de manera que se promueva la función reguladora por medio de acciones mediante las cuales el estudiantado exprese su punto de vista, ayude a sus compañeros(as) a superar dificultades, negocie y reconstruya significados.

En la búsqueda de una evaluación formativa, se propone la evaluación alternativa, que permita generar un trabajo participativo, de interacción permanente que privilegia la participación del alumnado, a partir de la orientación y guía de la persona docente. La evaluación formativa promueve el aprendizaje significativo y auto gestionado por quien aprende, a partir de su propia reflexión y análisis, pero también del trabajo colaborativo entre los diferentes actores de la acción formativa.

Según Álvarez (2001), cuando se busca que la evaluación contribuya a un proceso formativo, los criterios de valoración y de corrección han de ser explícitos, públicos, publicados y negociados entre el profesor y los alumnos. Esto le dará al proceso, además de una intención formadora, un carácter de mayor transparencia, ecuanimidad y equidad.

## Fundamentos de la evaluación alternativa

En muchas ocasiones las propuestas evaluativas tradicionales han obedecido a una serie de actividades desarticuladas y aisladas con poco significado para el aprendizaje, pero de mucha importancia para los juicios y las sumas que tienden a discriminar a quienes no logran conseguir los puntajes necesarios para alcanzar cierto estatus como estudiantes. A diferencia de este tipo de evaluación a la que nos hemos habituado, se propone una evaluación alternativa, la cual toma como punto de partida la evaluación formativa y compartida y se propone como un continuum, como una acción procesual e integradora y que se plantea desde el diseño mismo del proceso de aprendizaje.

Los contenidos y procedimientos seleccionados para evaluar con finalidades calificadoras y los criterios de evaluación aplicados condicionan totalmente cómo el profesor enseña y cómo el alumnado estudia y aprende. Por ejemplo, si las preguntas de los exámenes son memorísticas y muy reproductivas de lo que dice el libro de texto o la unidad didáctica, es normal que el alumnado crea que solo es necesario estudiar el día antes del examen para después olvidarlo todo, y que no vale la pena trabajar de manera continuada.

Toda actividad y toda práctica de evaluación educativa que no forme, que no eduque y de la cual los sujetos no aprendan, debe separarse de los procesos de aprendizaje. La evaluación alternativa propone buscar formas osadas y originales de evaluar, que sean congruentes con los contenidos por aprender, que sean altamente significativas y que, a la vez, promuevan nuevas formas de aprender. Por lo tanto, precisamos innovar con formas distintas de evaluar, que vayan más allá de las habituales: los exámenes, las pruebas cortas, los cuestionarios con puntajes establecidos y respuestas predeterminadas.

### **3.24 La evaluación en el contexto de la metodología del Aprendizaje Basado en Problemas**

La evaluación debe ser un método más de enseñanza, y también una manera real y directa de posibilitar el aprendizaje.

En el ABP, la evaluación se constituye en una herramienta por medio de la cual se le otorga al estudiante la responsabilidad de evaluar su proceso de aprendizaje y de formación. Es decir, se

trata de visualizar la evaluación como un proceso en el que la responsabilidad es compartida por los estudiantes y los tutores.

Esta práctica comporta un cambio sustancial en relación con el método tradicional. El valor de la evaluación en el ABP es el de contemplar la evaluación individualizada, cualitativa y formativa. El estudiante tiene la posibilidad de evaluarse a sí mismo, a sus compañeros, al tutor, al proceso de trabajo en equipo y a los resultados del proceso.

La evaluación formativa, como columna vertebral de una docencia centrada en el estudiante, comprende la autoevaluación, la de sus pares y la del tutor. La autoevaluación, como procedimiento valorativo, le permite evaluar, orientar, formar y confirmar el nivel de aprendizaje de cada una de las unidades. Además, le proporciona ayudas para descubrir sus necesidades, la cantidad y la calidad de su aprendizaje, las causas de sus problemas, las dificultades y los éxitos en el estudio. La evaluación por pares le permite al estudiante recibir retroinformación sobre lo que es incapaz de conocer sin la visión de los demás (la parte desconocida por el yo y conocida por los otros de la Ventana de Johan) I I. La evaluación del tutor le permite a éste emitir juicios acerca del nivel alcanzado y de la calidad del aprendizaje logrado. Por otro lado, la evaluación sumativa o de certificación debe tener en cuenta los conocimientos que el estudiante ha adquirido, las habilidades que ha desarrollado y las actitudes que ha modelado. Esta evaluación sirve para justificar las decisiones académicas respecto a las calificaciones.

### **3.25 Elementos críticos del ABP**

Un reto o pregunta que desafía (y estimula)

El corazón de un proyecto -“de qué se trata”, si queremos resumirlo así- es siempre un problema que resolver o investigar o una pregunta que analizar y contestar. Puede ser algo muy concreto (“Vamos a organizar una feria de la salud en nuestro centro”) o muy abstracto (“¿Cómo son los intereses de poder en el mundo actual”).

Una pregunta interesante y motivadora hace que los alumnos vean mucho más sentido en lo que están aprendiendo. Ya no estudian para recordar conocimientos sino que aprenden porque

necesitan nuevos conocimientos para resolver problemas o responder a preguntas que les interesan y les afectan.

El problema o la pregunta desafiará a los estudiantes sin intimidarlos. Cuando asesoramos a profesores, les aconsejamos diseñar la pregunta o desafío (ellos solos o con los alumnos) en forma de una propuesta abierta y accesible que concreta el objetivo final del proyecto. Por ejemplo, siguiendo con los ejemplos anteriores “¿Cuáles son las principales amenazas para nuestra salud? ¿Cómo podemos combatir las?” o “¿El mundo es un tablero con el que alguien juega?”.

### Investigación en profundidad

Investigar es un proceso mucho más complejo que buscar información en un libro o en Internet. El proceso de investigación ocupa tiempo. Esto implica que un Proyecto de calidad debe ocupar varias semanas.

En ABP, la investigación es un proceso cíclico y reiterativo: los estudiantes se hacen preguntas, buscan y encuentran recursos con información, responden a las preguntas iniciales, se plantean preguntas aún más complejas... El proceso se repite una y otra vez hasta llegar a una solución o respuesta final completa y satisfactoria.

Los proyectos pueden utilizar diferentes fuentes de información, mezclando la tradicional búsqueda en libros o Internet con otras posibilidades ligadas al mundo actual: entrevistas con expertos, audición de conferencias... Más aún, los alumnos que estén creando un producto pueden analizar las necesidades y demandas de los potenciales destinatarios de sus textos y presentaciones (sus padres, sus compañeros, sus vecinos...) o incluso de hipotéticos clientes.

### Autenticidad

Cuando afirmamos de algo que es “auténtico” queremos decir que es “genuino” o “real”, que no es “falso”. En educación, el término “auténtico” hace referencia a la vinculación de lo que se está aprendiendo con el mundo real. Hay muchas formas de que un proyecto sea auténtico. La mayor parte de las veces será necesario combinar varias de las siguientes opciones:

Podemos situar el aprendizaje en un contexto real, enfrentando a los alumnos a problemas que las personas encuentran en su vida personal y/o profesional (por ejemplo, empresarios que tienen que diseñar un plan de negocios, concejales que deben decidir las políticas municipales, familias que tiene que buscar vías para ahorrar dinero...).

Otra posibilidad es que el proyecto implique que los equipos de aula usen recursos, criterios de evaluación, herramientas e incluso personas del mundo real. Ejemplos muy sencillos de estos son aquellas tareas en las que los alumnos tienen que editar recursos como presentaciones, vídeos y audios o deben hacer una investigación en la que consultan a expertos en el tema.

Una tercera forma de vincular los proyectos a la realidad es que el resultado afecte positivamente a otros. Por ejemplo, si planificamos y construimos juguetes para el aula Infantil, estaremos desarrollando un proyecto en el que directamente estamos cambiando y mejorando las condiciones del aula de nuestros compañeros.

Por último, un proyecto será “auténtico” cuando se centra en los intereses, expectativas, conocimientos y contexto de los alumnos.

#### Decisiones de los alumnos

Dar la palabra (y la decisión) a los estudiantes provoca que estos sientan como propia la propuesta y trabajen y se involucren más en él. Si no permitimos que nuestros alumnos tomen decisiones acabarán viendo el proyecto como otro ejercicio o tarea de clase cerrado y pautado.

Evidentemente, el profesor tiene que aportar orientaciones y un cierto grado de control a sus alumnos en aspectos como las preguntas y respuestas que generan, los recursos informativos que utilizan, el reparto de tareas en los equipos de aula o los productos finales que generan.

Algunos grupos pueden dar uno o varios pasos más, decidiendo el tema y objetivo del proyecto, y redactando la pregunta inicial y decidiendo cómo van a investigar, cómo van a demostrar todo lo que han aprendido y finalmente cómo van a compartir el resultado de su aprendizaje.

#### Reflexión



John Dewey, cuyos planteamientos siguen conformando nuestra manera de enfocar el aprendizaje por proyectos dijo “No aprendemos de la experiencia sino reflexionando acerca de la experiencia”. A lo largo del proyecto, los estudiantes y el docente deben reflexionar acerca de qué están aprendiendo, cómo están aprendiendo y por qué están aprendiendo.

La reflexión sobre el aprendizaje puede tener lugar de manera espontánea e informal, como parte de las dinámicas de aula. No obstante, dentro del trabajo por proyectos tiene que ser diseñada sistemáticamente mediante diarios de aprendizaje, evaluaciones programadas, hitos de reflexión y revisión y presentaciones públicas de los trabajos y actividades de los alumnos.

Revisar los conocimientos adquiridos ayuda a los estudiantes a consolidar lo que han aprendido y pensar cómo pueden aplicar esos conocimientos en el proyecto. Reflexionar acerca de las competencias adquiridas permite que los alumnos tengan claro para qué sirven esas competencias y cómo pueden ayudarles en el futuro.

Revisar y evaluar el propio proyecto (documentos de reflexión y revisión del proyecto) ayuda a que los alumnos decidan cómo deben enfocar los siguientes proyectos y facilita que los docentes mejoren su práctica de aula y el diseño de las experiencias didácticas ABP.

### Crítica y revisión

Uno de los distintivos del aprendizaje por proyectos es la alta calidad de todas las actividades y tareas de aula de los alumnos. Esta calidad es conseguida mediante la revisión y el análisis crítico. Los estudiantes tienen que aprender cómo hacer evaluación entre iguales para dar y recibir críticas y análisis para que permitan mejorar sus proyectos y los de sus compañeros. Este análisis de los proyectos debe estar guiado por rúbricas, modelos y dinámicas de aula. [Enlace a recursos Slide share o artículo.](#)

Como complemento al análisis de profesores y compañeros, adultos del ámbito más cercano y expertos en los diferentes temas pueden contribuir para analizar el proyecto desde el punto de vista del mundo real.

La importancia que el sentido común da a la necesidad de revisar para mejorar continuamente los proyectos está respaldada por las investigaciones acerca de la evaluación formativa, que

implica no solo que profesores y expertos den a los estudiantes retroalimentación acerca de sus actividades sino que los alumnos evalúen su aprendizaje.

### Producto final público

Hay tres razones fundamentales para crear un producto final, que puede ser algo concreto y tangible o algo como una presentación pública o la respuesta a un problema planteado:

En primer lugar; un producto público acrecienta, al igual que la autenticidad, enormemente la capacidad de motivación de una propuesta didáctica ABP al mismo tiempo que refuerza en los alumnos la necesidad de llevar a cabo un trabajo de aprendizaje de gran calidad. Ejemplo: “La curiosidad es saludable”.

Pensemos por ejemplo en las presentaciones que los alumnos hacen en el aula para su profesor y sus compañeros. El riesgo no es muy alto por lo que los estudiantes puede relajarse y no tomarse muy en serio la actividad, con lo que la calidad de la presentación es baja. Sin embargo, cuando la presentación se hace para personas de fuera del aula la autoexigencia es mucho mayor ya que nadie quiere hacer un mal papel en público. Ejemplo: Cuando vivían en blanco y negro

Un cierto nivel de ansiedad es un buen motivador. Una ansiedad excesiva puede entorpecer la presentación en público. La cuestión es encontrar el grado justo de preocupación por el trabajo y para ello los alumnos tienen que estar preparados para hacer públicos sus trabajos.

En segundo lugar, al crear un producto los alumnos convierten lo que han aprendido en algo concreto y en consecuencia, cuando lo comparten, en algo cuestionable, discutible. La dimensión social de aprender cobra importancia y el aprendizaje deja de ser un intercambio privado entre el docente y los estudiantes.

Esto tiene un impacto en la cultura y la dinámica del aula, que se convierte en una comunidad de aprendizaje. El docente y sus alumnos analizan y discuten qué se ha aprendido, cómo se ha aprendido, cuáles son los criterios y estándares de evaluación que deben usarse y cómo se puede mejorar el trabajo de los alumnos.

Por último, publicar los trabajos de los alumnos es una manera muy efectiva de hacer llegar a los padres, a la comunidad educativa y a todo el mundo qué es el aprendizaje por proyectos y cómo esta metodología ayuda al aprendizaje de los estudiantes. Cuando un grupo, un instituto o una escuela se abren a la mirada y la revisión pública, el mensaje que están lanzando es: “Esto es lo que nuestros alumnos pueden hacer. Nosotros no nos dedicamos solo a hacer exámenes”.

Muchos docentes y escuelas refuerzan este mensaje reformulando la idea de “escuela abierta” en el sentido de hacer públicos los trabajos de los alumnos. De esta forma se refuerza el apoyo y la comprensión acerca de la metodología ABP en los padres, responsables educativos y otras personas del ámbito del centro. Cuando las personas ven la calidad de los productos (recordemos, en sentido amplio) que los alumnos pueden crear, quedan absolutamente sorprendidas y demandan ver más.

### **3.26 La evaluación por portafolios**

Un portafolios es una colección de documentos del trabajo del estudiante que exhibe su esfuerzo, progreso y logros.

El portafolio es una forma de evaluación que permite monitorear el proceso de aprendizaje por el profesor y por el mismo estudiante, permite ir introduciendo cambios durante dicho proceso.

Es una forma de recopilar la información que demuestra las habilidades y logros de los estudiantes, cómo piensa, cómo cuestiona, analiza, sintetiza, produce o crea, y cómo interactúa (intelectual, emocional y socialmente) con otros, es decir, permite identificar los aprendizajes de conceptos, procedimientos y actitudes de los estudiantes. Puede utilizarse en forma de evaluación, co-evaluación y de autoevaluación.

Cómo se aplica:

El estudiante puede participar en la selección de los contenidos, de los criterios de selección, de los contenidos para juzgar sus méritos y de la evidencia de la autoreflexión.

Los portafolios permiten al alumno participar en la evaluación de su propio trabajo. Al profesor le permiten elaborar un registro sobre el progreso del estudiante y le da bases para evaluar la calidad de su trabajo o de su desempeño en general. Existen distintos tipos de Portafolios de acuerdo al soporte material que utilizan (electrónicos o físicos, estos son carpetas de argollas) y conforme a su contenido:

#### Los portafolios de habilidades

Los estudiantes necesitan ser guiados para que puedan progresar en el desarrollo de sus herramientas personales. Cuando el alumno inicia su portafolio al comienzo de la licenciatura, al finalizar sus estudios, podrá observar en forma completa cómo contribuyó su educación en su formación y reconocer las destrezas que ha desarrollado.

El Portafolios se divide en áreas de capacidades. Los estudiantes insertan en las diferentes secciones los documentos que demuestren el desarrollo de capacidades y destrezas específicas.

Portafolios para desarrollar los currículum personales o historias de vida.

Por lo general cuando terminan sus estudios, los alumnos no advierten las habilidades que desarrollaron. Obtienen un documento con su historia académica que muestra los cursos tomados y las calificaciones recibidas. Conocen las actividades extracurriculares en que han participado, pero no pueden traducir esta información a las necesidades con que se van a encontrar en su vida profesional.

Los portafolios ayudan a los alumnos a identificar las capacidades que han desarrollado y los aprendizajes que han construido y cómo pueden utilizarlos; a reconocer los que son significativos y en un futuro los capacita para mostrarlas en forma ordenada al solicitar empleo en el mundo laboral.

Un portafolio completo puede incluir numerosas constancias escolares: bitácora personal, reconocimientos, ejemplos del trabajo escolar; currículum personal, que incluya los cursos tomados y sus resultados en cometido o desempeños, que se comparan con las calificaciones recibidas en el transcurso del curso; evidencia de su capacidad académica, por ejemplo

incluyendo una carta de recomendación de un profesor, etc. El portafolio podrá organizarse de la misma manera en que presentará los documentos al enfrentarse al mercado de trabajo.

Como docente es posible hacerse algunas preguntas para definir el propósito: ¿Cuál es mi visión de logro para mis estudiantes? ¿Qué se debe incluir en un portafolio? ¿Cuándo? ¿Cómo deberá verse? ¿Qué preguntas debe responder el estudiante sobre los agregados seleccionados? ¿Qué preguntas debe hacer el estudiante? ¿Qué preguntas debe hacer el profesor? ¿Quién más debe estar involucrado en la evaluación del portafolio: otros estudiantes, otros profesores, algunos expertos? Respondiendo a estas preguntas se puede definir el propósito.

### Evidencias

Cada evidencia debe organizarse para demostrar su evolución hacia la meta propuesta. Los tipos de evidencias pueden ser:

**Armazón:** Documentos del trabajo normal de grupo, desde actividades de clase hasta trabajos.

Realizados por iniciativa propia.

**Reproducciones:** Incluyen hechos que normalmente no se recogen, por ejemplo, grabación de un invitado o algún experto en el área.

**Testimonios:** Documentos sobre el trabajo del estudiante preparado por otras personas, por ejemplo, comentarios llevados a cabo por personas involucradas en el proceso formativo del estudiante.

**Reflexiones del profesor:** Producciones elaboradas por el estudiante, en donde éste explícita las metas del portafolio, incluye las reflexiones que lleva a cabo mientras lo elabora, se organiza o se evalúa el portafolio. Los documentos deben ir acompañados por breves informes que expliquen qué son, por qué se agregaron y de qué son evidencia.

**Evidencia de un cambio conceptual.** Identificando las pistas de los cambios en las concepciones del alumno sobre las ideas que se han visto en clase. ¿Cuándo ocurrieron los cambios? ¿A qué atribuyes estos cambios? ¿Qué hiciste para que ocurrieran?

Evidencia de crecimiento o desarrollo. Son una serie de trabajos que se guardan ordenados de tal manera que permitan observar la secuencia del aprendizaje. ¿Qué aprendiste? ¿Cómo?

Evidencia de reflexión. El alumno puede hacerse una secuencia de preguntas: ¿En qué son similares o diferentes? ¿Cómo se alteró mi percepción? ¿Cómo cambió mi comprensión?

Evidencia de toma de decisiones. Se presenta un ejemplo que demuestre la capacidad del alumno para advertir los factores que influyen en las decisiones que realiza. ¿Qué factores discutiste o pensaste? ¿Qué más necesitas saber para tomar diferentes decisiones? ¿Qué información requieres para apoyar las que ya realizaste?

Evidencia de crecimiento personal y comprensión, en donde el alumno se pregunte ¿Qué más sé? ¿Por qué? ¿Cómo?

### **3.27 La validez y fiabilidad en los enfoques alternativos de la medición y evaluación educativa**

La fiabilidad en psicometría se puede conceptualizar de distintas maneras que, a su vez, se traducen en métodos distintos de cálculo, los cuales son totalmente diferentes al proceso de validez. En ese sentido, cuando hablamos en términos propios de la fiabilidad de estas evaluaciones, nos referimos, sobre todo, a dos elementos: unanimidad y estabilidad:

La unanimidad indica directamente al punto en que el estudiantado, es consistente o unánime en su juicio cuando hace diferencias entre el profesorado. En otras palabras, pueden observar muy claro cada una de las características de uno u otro docente y de esa forma emitir un juicio global que se agrupa en la escala, no se dispersa la opinión hacia todos los puntos de ella, sino que se reúne alrededor de las características presentadas.

La estabilidad expresa el punto en el que las evaluaciones no varían notablemente al pasar del tiempo, sino que el cuerpo estudiantil mantiene los mismos criterios de valoración del profesorado con el transcurrir de los años. De esa forma, la evaluación del cuerpo docente, no cambiará mucho a través del tiempo y los resultados de su evaluación serán similares varios años después.

Estos dos aspectos de la evaluación de la docencia universitaria, unanimidad y estabilidad, son considerados por Overall y Marsh (1980), quienes en un estudio longitudinal muy importante, explican que la fiabilidad en las encuestas del estudiantado se entiende como el acuerdo relativo (unanimidad) entre las valoraciones de diferentes estudiantes dentro de la misma clase, bajo la asunción de que cualquier varianza específica del grupo estudiantil es aleatoria y debería ser considerada como varianza de error. Es estable a lo largo de un período de tiempo de varios años, separando los dos conjuntos de estas valoraciones, ello podrá ser incluido como la varianza sistemática cuando los coeficientes de estabilidad a largo plazo se basen sobre sus respuestas individuales.

No obstante, sobre este tema Cruse (1987) señala que los coeficientes de fiabilidad indican que el estudiantado puntúa de forma consistente, de la misma manera en ocasiones diferentes al profesorado, pero ello no significa que estos evalúan exactamente la docencia universitaria, pues a los estudiantes se les plantea un modelo tradicional de docente, el cual se ha ajustado a ciertos factores característicos del “buen profesor”. Este argumento es un tema debatido por mucho tiempo, principalmente, porque los rasgos encontrados por los investigadores pueden o no ser apropiados para un tipo de docencia específica, entonces si eso sucede se dejarían sin valorar otros rasgos especiales de la población docente universitaria.

El problema de todo este trabajo radica en que por la multidimensionalidad y complejidad no solo de la labor docente, sino de la misma universidad en su estructura, se hace difícil encontrar un modelo global que ajuste perfectamente a todas las características involucradas en el campo específico del saber universitario. Sin embargo, vale la pena resaltar que las investigaciones realizadas alrededor del mundo indican la existencia de ciertos factores característicos de la docencia universitaria que pueden ser usados para medirla en su totalidad.

En ese sentido, Marsh (1987) afirma que en los instrumentos bien contruidos, la consistencia interna es normalmente alta, aunque “proporciona una estimación inflada de la fiabilidad, porque ignora la proporción sustancial de error debido a la falta de acuerdo entre diferentes estudiantes y por lo tanto no deberían ser usados en general” (pág. 275); quizá puede ser apropiado utilizarlas para determinar hasta qué punto, las correlaciones entre facetas múltiples se han hecho tan grandes que las facetas separadas no pueden ser distinguidas, como con el

procedimiento multirasgo multimétodo. Por lo tanto, el autor recomienda que una de las medidas por utilizar es la fiabilidad de la respuesta media de clase basada en el acuerdo entre el estudiantado de la misma clase, o sea, la unanimidad.

### **3.28 La evaluación de referencia criterial versus la referencia normativa**

Evaluación Normativa:

Evaluar en referencia a una norma, significa comparar el resultado del individuo con los resultados de una población o grupo a los que pertenece.

Esto exige el establecimiento de una norma o escala de referencia, confeccionada después de estudios estadísticos de rendimiento, con el objetivo de obtener una calificación. En este ámbito normativo, el criterio es externo, en la medida que se utiliza una escala que es más o menos "ajena" al sujeto evaluado, sin tener en cuenta las condiciones de trabajo, nivel inicial, aprendizaje, etc. (B. Maccario). Por esta razón, este tipo de evaluación se utiliza para ubicar a los alumnos en escalas de rendimiento y puntaje, atribuir un lugar dentro de los grupos, certificar los niveles en función de la norma o el grupo y predecir futuros resultados.

Evaluación Criterial:

En la evaluación por criterios, los resultados de la medición se comparan con un criterio absoluto, constituido por lo que debe "saber hacer" el estudiante. A este "saber hacer" se le conoce como dominio, que implica tanto la clase de tarea que ha de ejecutarse, como el contenido implicado en la ejecución.

Evaluar en referencia a un criterio, busca la comparación del alumno con sus propios rendimientos o resultados, en las mismas pruebas o en relación a un criterio fijado de antemano. Se valora principalmente el progreso realizado por el alumno, independientemente de escalas, y se valora el proceso realizado por el alumno hacia el objetivo propuesto. En el ámbito criterial, se evalúa el avance del alumno hacia este objetivo y la distancia que lo separa de él. Esta distancia constituye las bases de la información a partir de la cual se ha de tomar una decisión. Esto nos aproxima a una "pedagogía por objetivos", donde existe una necesidad de expresar los objetivos en términos operativos (el alumno será capaz de...), luego de haber



analizado las necesidades y posibilidades del alumno o grupo. Las funciones que cumple este tipo de evaluación son: establecer un balance con los objetivos propuestos, realizar un diagnóstico de las dificultades y determinar si la estrategia es o no pertinente. Estos conceptos nos remiten a los fundamentos de la evaluación formativa.

En la formación profesional para el empleo, se trata de evaluar si los alumnos han alcanzado los objetivos previstos. Por tanto, creo que el tipo de evaluación que se utiliza es el criterial, ya que como se ha mencionado antes, en la evaluación criterial se valora principalmente el progreso realizado por el alumno, independientemente de escalas, y se valora el proceso realizado por el alumno hacia el objetivo propuesto. Esto a fin de determinar qué es lo que puede hacer el estudiante y si su ejecución es superior o inferior al dominio, independientemente de si es superior o no a la de los demás estudiantes; luego no se trata de comparar sus resultados con el grupo al que pertenece, si no de evaluar su “capacitación” en base a los criterios fijados por los objetivos a conseguir en la acción formativa.

### **3.29 Origen, desarrollo y naturaleza de las pruebas de referencia criterial**

Existe coincidencia en la literatura de evaluación a gran escala, en situar a Glaser como el autor del primer artículo publicado con el tema de evaluación referida a criterio, debido principalmente a que es el primero en plantear lo inadecuado de sustentar una variedad de decisiones acerca de programas instruccionales basados en objetivos o en competencias, a partir de mediciones obtenidas con las tradicionales pruebas normativas; el mismo Glaser, no obstante, señala que Flanagan en 1951 y Ebel en 1962 habían hecho una distinción entre evaluación normativa y criterial (Glaser, 1963). A partir de mediados de los años sesenta, se pueden encontrar en la literatura diversas definiciones de lo que es una prueba referida a criterio, entre las que destacan, por sus contribuciones al campo, la de Glaser y Nitko (1971), quienes enfatizan que una prueba criterial sirve para obtener mediciones directamente interpretables en términos de realizaciones estándar concretas, es decir, lo que el sustentante puede o no realizar; y la de Popham (1978), quien refería el uso de estas pruebas para determinar la posición de un individuo con respecto a un dominio perfectamente definido.

En cuanto a su uso, Hambleton y Swaminathan (1978) señalan que las evaluaciones referidas a criterio sirven para guiar el proceso individual en programas basados en objetivos de aprendizaje, comprobar el rendimiento de los alumnos, diagnosticar deficiencias de aprendizaje, evaluar programas educativos y de acción social y para verificar el logro de competencias con fines de certificación u otorgamiento de licencias. En este tipo de evaluación, la apreciación del grado con que un sustentante cumple con los objetivos de la enseñanza se lleva a cabo en función de su desempeño, sin compararle con el de sus compañeros.

Aunque existen algunas variantes en la definición del concepto de evaluación criterial, hay acuerdo en que sirve para comprobar el rendimiento mediante la apreciación de las realizaciones personales respecto de los objetivos logrados, sin compararlas con las del grupo al que pertenece, facilitando así el diagnóstico de dificultades, la programación de las actividades de recuperación y la toma de decisiones de promoción de nivel o de certificación de cada individuo evaluado.

Otros autores importantes que aportaron herramientas conceptuales al campo fueron Popham y Husek (1969), con una aclaración importante al destacar que el término criterio —de las pruebas criterioles— se refiere a un dominio de conductas bien definido y no sólo al hecho de haber establecido un estándar de ejecución o un punto de corte. En cuanto a la interpretación de la ejecución o desempeño de un individuo, es correcto hacer una interpretación de tipo descriptivo, o bien establecer algún estándar o punto de corte contra el cual contrastar dicha ejecución, lo cual será posible siempre que se haya definido adecuadamente el marco lógico contra el cual se va a contrastar tal ejecución o desempeño, es decir, que la adecuada definición del dominio resulta indispensable. Desde luego la decisión de establecer un punto de corte o hacer una interpretación descriptiva depende del propósito de la prueba, como se puede apreciar en las distintas pruebas que se aplican en la actualidad.

Una confusión frecuente en este campo es la relativa a las diversas denominaciones que se han empleado en la literatura para referirse a pruebas que presentan pocas o ninguna diferencia con las pruebas referidas a criterio (las denominadas pruebas de maestría, o las pruebas referidas a dominio, y las referidas a objetivos). Si se adopta la definición de Popham (1978), no existen diferencias esenciales entre ellas; finalmente todas se constituyen de reactivos que

emparejan con objetivos. La distinción principal entre pruebas referidas a objetivos y las referidas a criterio es que en estas últimas los ítems son un conjunto representativo de un dominio claramente definido de conductas que miden un objetivo, mientras que en las pruebas referidas a objetivos no siempre se especifica un dominio de conductas, por lo que no se considera que los ítems sean representativos de algún dominio conductual (Hambleton y Swaminathan, 1978).

Actualmente existen pocas dudas respecto de la necesidad de emplear pruebas con especificaciones cualitativamente diferentes a aquéllas que típicamente se han empleado para el diseño de una prueba referida a norma, y existe acuerdo en que las características más destacadas y comunes a lo que se ha dado en considerar una prueba referida a criterio son:

1. Requiere la definición clara y exhaustiva de un dominio objetivo a evaluar.
2. Permite averiguar la posición de un sujeto respecto del dominio de una conducta bien definida que manifieste el aprendizaje de un alumno.
3. La interpretación del rendimiento es directa: la ejecución que realiza el alumno indica su grado de dominio o competencia, independientemente de lo que hagan otros sujetos.
4. El criterio o estándar en el cual se basa tiene un carácter absoluto, es decir que no está condicionado por el nivel de ejecución de un grupo. Es la descripción de la clase de conducta que el alumno puede o no manifestar.
5. El límite en que se basa la toma de decisiones que afectan al proceso educativo se establece de manera descriptiva, indicando el grado de dominio alcanzado o bien especificando un punto que se toma como punto de corte, o nivel mínimo de dominio.
6. Permite retroalimentar la intervención en el proceso educativo de manera inmediata.

En términos metodológicos las pruebas referidas a criterio son especialmente útiles cuando se trabaja con unidades de aprendizaje muy concretas, es decir, aquéllas al servicio de la evaluación formativa (Jornet y Suárez, 1994); no obstante, en este artículo se destaca su utilidad para la evaluación sumativa aplicada a cursos escolares completos o incluso a periodos más largos

aplicada a grandes poblaciones, como es el caso de las pruebas a gran escala que se están aplicando actualmente en todo el mundo. Esta utilidad radica en su capacidad para:

1. Determinar el nivel de aprendizaje alcanzado por un alumno y, en función del mismo, asignarle una calificación o emitir un juicio de acreditación académica.
2. Certificar que un determinado alumno posee, al menos en el momento de la evaluación, ciertas destrezas, conocimientos, habilidades o competencias.
3. Predecir el éxito de un alumno en un curso posterior relacionado con la misma materia. Para que dicha predicción quede garantizada, el alumno al menos debe superar el punto de corte de aquellos objetivos que se consideran básicos, por su papel de conectores con los aprendizajes posteriores.
4. Indicar a los profesores, planeadores y autoridades educativas problemas de programación y secuenciación de áreas o temas en el currículo escolar.
5. Informar a los alumnos sobre su progreso en el aprendizaje, indicando los aspectos más logrados y los más deficientes.

#### Construcción de las pruebas criterioles a gran escala

Dado que las pruebas criterioles son construidas con la finalidad de apoyar generalizaciones válidas respecto del desempeño de los individuos en relación con dominios especificados de contenido o de conducta, Hambleton (1985) y Popham (1978) han enfatizado que los objetivos conductuales por sí mismos no son suficientes para una especificación detallada de los reactivos de la prueba. Las especificaciones de los reactivos, según Popham (1978), deben dividirse en cuatro partes: descripción de un objetivo conductual; dirección y ejemplos de los tipos de reactivos a incluir; acotación y limitaciones respecto de la amplitud y dificultad del contenido; y descripción del tipo de respuestas esperadas, así como de respuestas incorrectas. Con base en la investigación desarrollada en el campo durante más de dos décadas, Hambleton (1995) presenta una propuesta metodológica, que aún se considera vigente, de 12 pasos a seguir para la construcción de este tipo de pruebas:

1. Preparación y selección de las especificaciones del dominio o los objetivos que se pretenden evaluar.
2. Descripción clara y detallada de las especificaciones: los propósitos de la prueba, el tipo de formato de reactivos deseable, el número de reactivos de la prueba y las instrucciones para los redactores de reactivos.
3. Redacción de los reactivos para medir los objetivos incluidos en la prueba, o versiones de la prueba si se requiere la elaboración de formas paralelas.
4. Edición inicial de los reactivos de la prueba por los individuos que los redactan.
5. Evaluación sistemática y consistente de reactivos en los pasos 2 y 3 para determinar su congruencia con los objetivos y para determinar su representatividad.
6. Edición adicional de reactivos con base en los datos del paso 5, descartando aquéllos que no midan adecuadamente los objetivos que se pretenden medir.
7. Integración de versiones de la prueba con base en las especificaciones realizadas.
8. Empleo de métodos para determinar estándares (puntos de corte) que permitan interpretar la ejecución de los sustentantes.
9. Administración de la prueba bajo condiciones estandarizadas de aplicación.
10. Investigación y análisis para recopilar evidencias necesarias de confiabilidad y validez de la prueba.
11. Preparación de un manual técnico de la prueba.
12. Compilación de datos técnicos, tanto de los reactivos como de la prueba, para reforzar la validez en términos de las inferencias que se llevan a cabo y el tipo de usos en condiciones diversas y con diferentes poblaciones examinadas.

Resulta conveniente destacar que las operaciones fundamentales para la elaboración de este tipo de pruebas son: la especificación del dominio, el análisis de los reactivos, la determinación de estándares o puntos de corte, y la determinación de la confiabilidad y la validez; y que éstos son procesos iterativos hasta lograr niveles ajustados y satisfactorios. La definición del dominio

o universo de medida es origen y referencia de todos los demás, por lo cual debe reunir características de precisión tales, que en todo momento permita saber si un reactivo pertenece o no a dicho dominio; el punto de corte o estándar se entiende como el punto de superación o nivel mínimo, es decir, el valor que, dentro de un continuo de medida sobre el que se puede situar la ejecución de un individuo, sirve para diferenciar el grado de suficiencia del grado de insuficiencia en el cumplimiento de objetivos o el logro de competencias.

### **3.30 Características técnicas de las pruebas de referencia criterial, elementos diferenciales**

#### Confiabilidad (Fiabilidad)

Se refiere al hecho de la precisión de la medida, independientemente de los aspectos que se pretenden medir. Es la capacidad de un test para demostrar estabilidad y consistencia en sus resultados. Un test es confiable cuando al aplicarlo dos o más veces a un individuo, en circunstancias similares obtenemos resultados análogos o similares. La confiabilidad debe controlar la distorsión o variación que diversos factores producen al aplicarlo, alguno de ellos pueden ser:

- El clima
- La hora del día
- Las instalaciones
- El equipo de medición
- El estado de ánimo del evaluado.

Para calcular la confiabilidad de un test podemos aplicar el procedimiento de "test-retest", donde la misma prueba es aplicada a un grupo, en dos oportunidades, luego de un corto intervalo, y el coeficiente de correlación obtenido nos demuestra la seguridad y precisión del test. (Coeficiente de confiabilidad)

#### Objetividad

Una prueba de evaluación (test) es objetiva cuando sus resultados son independientes de la actitud o apreciación personal del observador. Es el grado de uniformidad con que varios

individuos pueden aplicar un mismo test. Para ello se buscan instrumentos precisos, tiempo - cronómetro, distancia - cinta métrica, etc. La objetividad garantiza la confiabilidad de un test.

La objetividad debe observarse en:

Objetividad de realización, en la construcción, aplicación, explicación, descripción e instrucciones del test.

Objetividad de evaluación, puede ser métrica (sistema internacional de medidas) o calificadora (evaluación subjetiva)

Objetividad de interpretación, los grados de valoración del rendimiento obtenidos en las pruebas dependen también si la prueba es en si misma objetiva o cuando incide más el margen de interpretación, o sea en las pruebas subjetivas.

Cuando al menos 2 evaluadores utilizan el mismo test con el mismo grupo y en intervalos reducidos, podemos calcular el coeficiente de correlación de objetividad. El test será objetivo si los resultados obtenidos por ambos evaluadores son análogos, mostrando una  $r$  significativa.

Estos criterios de calidad principales para la aplicación de un test (validez, confiabilidad y objetividad) se acompañan de una serie de criterios secundarios, pero de gran importancia para su aplicación práctica. (M. Grosser)

### Normalización

Consiste en la transformación del valor del test (intangible) en una ubicación con relación a una norma. Esto lleva a la realización de escalas de medida, que se elaboran a través de estudios estadísticos poblacionales, con el objetivo de la confección de las normas.

### Estandarización

Para que sea válida la comparación de resultados recogidos sobre diferentes grupos, o sobre el mismo grupo en períodos diferentes, es necesario uniformizar (estandarizar) las técnicas de administración de los test. Una pequeña variación en las normas de realización de una prueba puede alterar el resultado y su valoración posterior.

### Economización

Se considerará una prueba económica aquella que es realizable en poco tiempo, que precisa poco material y aparatos, que es fácil de manejar y que puede ser interpretada fácilmente sin muchos cálculos.

#### Utilidad

Se considera útil un test que analiza una conducta o capacidad para cuyo conocimiento hay una necesidad práctica y un auténtico interés de conocimiento.

#### Probabilidad

El test debe poder ser realizado con éxito por el 90% de los alumnos, que hayan seguido regularmente el proceso de aprendizaje. (A.P. Teleña)

### **3.31 La evaluación referida a estándares**

A diferencia de las evaluaciones más tradicionales, las evaluaciones impulsadas por estándares están estrechamente vinculadas al currículo, produciendo una conexión estrecha entre lo que se enseña y lo que se evalúa. A diferencia de las pruebas que referidas a normas, que comparan el desempeño de cada estudiante con el de otros, las evaluaciones impulsadas por estándares incorporan metas de desempeño previamente establecidas. Y, a diferencia de los exámenes de opción múltiple, muchas evaluaciones impulsadas por estándares requieren que los estudiantes demuestren una amplia gama de habilidades para la solución de problemas -- las mismas habilidades que los estudiantes necesitarán para su éxito futuro.

Estas evaluaciones “auténticas” o evaluaciones “de desempeño” suelen hacer que los estudiantes se involucren en problemas de la vida real más que en ejercicios artificiales. No sólo miden la habilidad que tienen los estudiantes para manejar tareas complejas, sino que también sirven de modelo para los maestros, proporcionándoles ejemplos a ser usados en el salón de clases.

Las evaluaciones de desempeño que requieren respuestas extensas deben ser calificadas por jueces expertos, usando guías de calificación claramente especificadas. El desarrollo y la aplicación de dichas guías de calificación brindan a los maestros una oportunidad inusitada de discutir estándares nuevos y expectativas de desempeño.



Examinar las respuestas que dan sus alumnos permite a los maestros comprender los puntos fuertes y las debilidades del aprendizaje de sus pupilos y planear actividades instruccionales apropiadas.

¿Qué es lo que hace sólida a una evaluación? Típicamente, se citan dos criterios principales: validez, el grado hasta el cual se justifican usos e interpretaciones particulares de los resultados de la evaluación; y confiabilidad, el grado hasta el cual las calificaciones están libres de error de medición. Para la evaluación impulsada por estándares, otra clave es el alineamiento -- el grado hasta el cual la evaluación refleja adecuadamente los estándares en los cuales se supone está basada. Una evaluación mal emparejada con un conjunto dado de estándares puede minar el aprendizaje al enfocar la atención en habilidades o conocimientos menos importantes a costa de otros más importantes.

Retos para los sistemas de evaluación impulsada por estándares

Construir un consenso estatal y local

Si nos guiamos por las encuestas de opinión pública, el concepto de que los estudiantes deben alcanzar estándares académicos altos goza de amplio respaldo. Sin embargo, la experiencia muestra que dicho respaldo puede ser frágil. La diversidad de opiniones sobre qué es lo que los estudiantes deben aprender y qué es lo que las escuelas deben enseñar hace imperativo el involucrar al público en el desarrollo de estándares y evaluaciones. Construir un consenso amplio requiere no sólo una serie de audiencias públicas y oportunidades para hacer sugerencias y revisiones, sino un proceso integral que involucre intensamente al público, asegurando que sus preocupaciones sean comprendidas y atendidas.

Proporcionar estándares fuertes

Lograr un consenso sobre estándares que son vagos y amplios no es ningún reto -- ¿quién estaría en desacuerdo con que todos los estudiantes deben ser capaces de “comunicarse efectivamente?” Pero cuando los estándares están formulados en términos tan generales, son de poca ayuda para los estudiantes que los deben lograr o para los maestros y escuelas que intentan evaluar el progreso estudiantil. La evidencia disponible sugiere que los estándares

actuales de muchos estados no son lo suficientemente fuertes como para sustentar una evaluación rigurosa basada en contenidos.

#### Alinear los estándares con la evaluación y la enseñanza

Muchos estados y localidades desarrollan estándares y evaluaciones al mismo tiempo, en lugar de seguir la secuencia más lógica: primero los estándares, luego las evaluaciones. De hecho, algunos estados construyen sistemas de evaluación en base a “retazos” de cuantas evaluaciones y pruebas puedan encontrar, sacrificando el “entalle perfecto” que obtendrían si desarrollaran evaluaciones empezando por el principio. Los sistemas que se basan exclusivamente en exámenes de opción múltiple no pueden mostrar qué tan bien se están desempeñando los estudiantes en la gama completa de habilidades y comprensiones que abarcan los estándares.

Por último, el currículo y la instrucción en el aula también deberían estar alineados con los estándares y las evaluaciones. Sin embargo, este alineamiento depende a su vez de la capacidad de los maestros para comprender las expectativas incorporadas en las nuevas evaluaciones para obtener recursos y experiencia para ayudar a sus alumnos a alcanzarlas. Es cuestión de justicia que los estudiantes no sean responsabilizados por metas que no han tenido oportunidades adecuadas de alcanzar.

#### Asegurar mediciones precisas

Las evaluaciones de desempeño, que piden a los estudiantes generar una respuesta en lugar de elegir una de una lista, generalmente proporcionan un mejor indicador de habilidades para el pensamiento complejo. Pero calificar evaluaciones como esas requiere más tiempo, generalmente más dinero y un consenso entre jueces sobre la calidad de la respuesta. Para proporcionar un estimado estable sobre la capacidad estudiantil, la mayoría de las evaluaciones que ahora se desarrollan incorpora una variedad amplia de tareas, reflejando así el alcance completo de los estándares. Cuando se mide el desempeño o progreso de una escuela o distrito, en lugar del de un estudiante individual, las evaluaciones pueden también asignar tareas diferentes a muestras diferentes de estudiantes (una práctica conocida como muestra matricial).

### Definir el progreso

El progreso de escuelas, distritos y estados se define típicamente por el desempeño de cohortes sucesivas de estudiantes: ¿Están, por ejemplo, más chicos de 4° grado demostrando su proficiencia en estándares matemáticos este año que el año pasado? La ley federal requiere que los estados definan un “progreso anual adecuado” en términos del desempeño estudiantil en las evaluaciones estatales impulsadas por estándares, que determinen si sus escuelas están logrando dicho progreso, y que señalen como meta una fecha “apropiada” para la cual todos sus estudiantes del Título I se desempeñarán ya sea en el nivel proficiente o en el avanzado. Los estados deben luego establecer una tasa anual de mejoría que sea tanto “sustantiva” como “suficiente” para lograr esa meta.

### Establecer las implicancias (stakes)

Lo que las escuelas hacen con los resultados de las evaluaciones -- ya sea simplemente reportándolas, en un extremo del espectro, o condicionando la graduación de los estudiantes, al otro extremo -- puede tener efectos profundos sobre esos estudiantes. Las evaluaciones también pueden utilizarse para responsabilizar a los educadores y a las escuelas por el desempeño de los estudiantes. Los distritos pueden utilizar los resultados como base para premios explícitos (otorgamiento de donaciones en efectivo, por ejemplo) o sanciones (reassignación o destitución de personal, reorganización administrativa).

### Incluir a todos los estudiantes

Los estándares están diseñados para elevar las expectativas que se tienen para todos los estudiantes. El incluir en una evaluación a estudiantes con limitaciones en la lengua inglesa y a estudiantes discapacitados puede requerir una variedad de distintas estrategias para brindarles facilidades, desde la asignación de tiempo extra hasta la provisión de evaluaciones orales o la traducción a otras lenguas. Los estudiantes con problemas de aprendizaje, que históricamente ha constituido el grupo más grande excluido de las evaluaciones, pueden ser capaces de rendirlas, ya sea en parte o por completo, sin facilidades especiales. (Aquéllos con problemas cognitivos severos pueden requerir de un sistema de evaluación distinto).

Estimar costos

Aunque los costos de las evaluaciones varían considerablemente, aquéllas que requieren respuestas extensas de los estudiantes -- a ser evaluadas por maestros u otros expertos en las disciplinas -- cuestan sustancialmente más que las pruebas de opción múltiple.

### **3.32 La evaluación basada en ejecuciones**

Conjunto de pruebas en las cuales a un individuo se le pide que realice una tarea específica.

Evaluada de manera objetiva, sistematizada y de manera uniforme.

Se usan criterios predeterminados como el número de repeticiones o el tiempo de duración de la actividad para su calificación.

#### **VENTAJAS**

Observación directa; objetiva y cuantificable

Elimina discrepancias en la evaluación entre los ancianos y sus cuidadores

Predicen eventos como Caídas, Admisión a hogares, Mortalidad

Requieren un mínimo de función cognoscitiva disminuyendo sesgos en la evaluación de los ancianos con deterioro.

Individualiza la discapacidad o la limitación funcional, así permite hacer un enfoque de prevención más adecuado a las necesidades del paciente.

#### **DESVENTAJAS**

Entrenamiento mínimo

Fallas en evaluar la actividad exacta

Fatiga, cansancio o no deseo de realizarla.

Algunas pueden conllevar a un riesgo, como pararse en un solo pie.

## UNIDAD IV

### LAS COMPETENCIAS Y SU EVALUACION

#### 4.16 Introducción

La educación por competencias ha resurgido en todo el mundo con gran fuerza desde finales del siglo pasado. Las reformas del sistema educativo mexicano emprendidas en la última década y que comprenden todos los niveles tienen un elemento en común: un currículo con un enfoque basado en competencias. Al igual que en otros países, en México estamos viviendo una efervescencia por las competencias; no obstante, la confusión y la incertidumbre parecen ser el signo distintivo de estas reformas. Es cierto que el concepto de competencias es polisémico y complejo, que no existe una teoría unificada que sustente este enfoque y que carecemos de experiencia en la puesta en práctica de un currículo basado en competencias. Si a esto le sumamos la dificultad propia que entraña la evaluación del aprendizaje, diremos que las condiciones están puestas para que la ambigüedad y el desconcierto reinen en los centros educativos, tal como está sucediendo. Este artículo tiene como objetivo contribuir al debate teórico actual en torno a la evaluación de competencias, así como aportar algunos rudimentos de tipo metodológico que permitan a los profesores mejorar sus concepciones y prácticas de evaluación.

#### 4.17 Conceptualizando el termino competencia

El concepto de competencia es, probablemente, uno de los términos más referidos en la conformación de los discursos pedagógicos actuales. Su empleo y uso variado, en diversos contextos, ha permitido que se identifiquen como competencias realidades ajenas a esta o, en el mejor de los casos, que se confunda por competencia términos que pese a ser afines no cumplen los atributos sustanciales de la propia competencia.

La competencia es un concepto polisémico y difícil de aprehender; existen diversas formas de concebirlo y el consenso parece una lejana quimera. En la bibliografía se pueden encontrar distintas posturas sobre las competencias, algunas francamente pobres y con escasa fundamentación conceptual (por ejemplo, los trabajos de Tobón y de Frade), alimentadas por

una visión en exceso pragmática de las competencias, salpicadas de conceptos de distinta procedencia teórica (conductismo, constructivismo, cognitivismo, humanismo, paradigma de la complejidad, enfoques empresariales...) yuxtapuestos sin rigor ni consistencia, las cuales gozan de una inmerecida popularidad entre el profesorado del país; quizá en su simplicidad y excesiva superficialidad radica su éxito; hasta enfoques mucho más amplios y comprensivos que abordan las competencias como un objeto complejo, dinámico y multidimensional orientados por una perspectiva constructivista sociocultural (Escudero, 2008; Perrenoud, 2008; De Ketele, 2008; Tardif, 2008; Bolívar, 2008 y 2010; Roegiers, 2010).

#### **4.18 Conocimiento, capacidad, habilidad, competencia**

La capacidad, la habilidad y la competencia son aptitudes, que todos los seres humanos tenemos por naturaleza y, aunque todas estas aptitudes están ligadas son muy diferentes pero, es muy fácil confundir un término con otro.

**CAPACIDAD:** Del latín “capacitas”, es la aptitud con que cuenta cualquier persona para llevar a cabo cierta tarea. Es decir que todos los seres humanos estamos capacitados para realizar con éxito cualquier tarea, Si bien este término se confunde con **TALENTO**, por lo cual es necesario aclarar las Diferencias entre ambos.

La capacidad es una serie de herramientas naturales con las que cuentan todos los seres humanos. Se define como un proceso a través del cual todos los seres humanos reunimos las condiciones para aprender y cultivar distintos campos del conocimiento, aun si estas condiciones hayan sido o no utilizadas, de esta manera, nos referimos a estas condiciones como un espacio disponible para acumular y desarrollar naturalmente conceptos y Habilidades.

Es muy común hablar de capacidad, sin embargo este término debería usarse para describir la **FLEXIBILIDAD MENTAL DE UNA PERSONA**, que sirve para medir con cuanta facilidad una persona se enfrenta a un nuevo reto intelectual y es aquí donde se confunde capacidad con el talento.

De esta manera el **TALENTO** es una actividad o área específica del conocimiento en particular, aunque, también se puede referir a esa porción de la capacidad de una persona que sirve específicamente para interiorizar un concepto o una disciplina, sin embargo es importante

enfaticar que el talento está formado por un conjunto de habilidades naturales, pues en muchas ocasiones el talento se consigue en base a la dedicación para aprender y mejorar, el resultado de cualquier tarea y luego de un trabajo intenso.

En el tiempo actual, los seres humanos desperdiciamos la mayor parte de nuestra capacidad mental, por esta razón muchas veces nos sucede que descubrimos habilidades que jamás habíamos explorado antes, y la educación es el principio en la aventura que representa aprender y especializarse, para conocer y entender que nuestros límites en cuanto a la capacidad están mucho más lejos de lo que pensábamos.

**HABILIDAD:** Del latín “habilitas”, hace referencia a la maña, la facilidad, aptitud y rapidez para llevar a cabo cualquier tarea o actividad.

Por lo tanto una persona hábil, es aquella que obtiene éxito gracias a su destreza.

Habilidades; Quién ha desarrollado, habilidad, "ha adquirido una metodología, técnicas específicas y práctica en su área de trabajo. Prácticas que han de abreviar formas de procesos intelectuales o mentales

Las personas tenemos diferentes tipos de conocimientos y diferentes competencias; pero solo un grupo de ambos se pone en acción cuando hacemos algo. Cuando se hace referencia la habilidad de un integrante en un equipo, solo se piensa en el talento en relación con la tarea a realizar, sin que signifique que esta habilidad lo tenga en la realización de otra tarea.

En el desarrollo personal la habilidad es un cierto nivel de competencia para cumplir con una meta específica.

Habilidad cognitiva: personas que son capaces de solucionar diferentes problemas, de comprender que una determinada acción tendrá una serie de consecuencias y por lo cual tomar una serie de decisiones

No obstante, no podemos pasar por alto el hecho de que, además de este tipo de habilidad, existen otras clasificaciones que permiten determinar las distintas variantes de aquella.

Habilidad matemática: inteligencia que emplean las personas y que se utilizan en forma correcta para completar un proceso de raciocinio de manera adecuada. Por lo tanto, las personas con

habilidad matemática, tienen facilidad para trabajar con funciones, proporciones y otros elementos abstractos.

Habilidad social: todas aquellas capacidades que tiene una persona para llevar a cabo lo que son las relaciones de tipo interpersonal. Así, se hace referencia a la capacidad para comunicarse, para tener empatía o para negociar.

Las habilidades de tipo comunicativo son las terceras más importantes y son aquellas que determinan la capacidad de alguien para llevar a cabo el análisis de la influencia que puede tener con otras personas o incluso los medios de comunicación. Sin olvidar que también hará lo mismo con los valores o las normas que estén establecidas en la sociedad.

Habilidades para el trabajo. En concreto, las que son más valoradas en un profesional son el liderazgo, la creatividad, la destreza manual, la capacidad lingüística o la destreza de tipo mecánico. Todas ellas determinarán que aquel sea un empleado competente, eficaz e importante para cualquier empresa.

Cabe destacar que la habilidad puede ser una aptitud innata (es decir, transmitida por la vía genética) o desarrollada (adquirida mediante el entrenamiento y la práctica). Por lo general, ambas cuestiones se complementan: una persona puede haber nacido con habilidad para jugar al fútbol, pero tendrá que entrenar muy fuerte si quiere desarrollar su talento y poder competir a nivel profesional.

#### **4.19 Tipología de las competencias**

Saber sumar, restar, multiplicar o dividir ya no es suficiente para que un alumno de primaria supere con éxito una evaluación matemática. Si este alumno no es capaz de aplicar estas operaciones a un contexto real de la vida cotidiana, no habrá conseguido desarrollar una de las competencias básicas a las que se orienta el aprendizaje de tal área de formación y por tanto no habrá alcanzado los objetivos de la asignatura.

A partir de ahora, tanto los objetivos de cada área de aprendizaje o materia, como los criterios de evaluación, la organización del colegio, la participación de las familias... todos los aspectos formales y no formales afectan a la educación de los jóvenes.



Generalmente las escuelas diseñan los currículos de primaria y secundaria en función de estas competencias:

- Competencia en comunicación lingüística.
- Competencia matemática.
- Competencia en el conocimiento y la interacción con el mundo físico.
- Tratamiento de la información y competencia digital.
- Competencia social y ciudadana.
- Competencia cultural y artística.
- Competencia para aprender a aprender.
- Autonomía e iniciativa personal.

¿Por qué se han introducido estas competencias en el currículo?

Principalmente, como una respuesta a la nueva demanda en materia de educación que requiere la sociedad actual. Es decir, una educación y formación que, más que enfocada a la pura adquisición de conocimientos, se oriente al desarrollo de destrezas y habilidades que resulten útiles para los jóvenes a la hora de desenvolverse de manera autónoma en la vida diaria.

Además de "saber", los alumnos deben saber aplicar los conocimientos en un contexto real, comprenderlos y tener la capacidad de integrar los distintos aprendizajes, relacionarlos entre sí y utilizarlos de manera práctica en las posibles situaciones o circunstancias a las que tengan que enfrentarse diariamente.

Asimismo, la labor del profesorado es fundamental para alcanzar los objetivos marcados por las competencias básicas que, además de los cambios implicados en el modo de enseñar, deberán evaluar a los alumnos no sólo por los conocimientos requeridos, sino en la medida en que éstos han contribuido a la adquisición de las competencias, y deben enfocar la acción tutorial a este objetivo, orientando y estimulando de manera personalizada el proceso de aprendizaje de los alumnos.

Las competencias

Competencia en comunicación lingüística:

La adquisición de esta competencia supone que el estudiante es capaz de utilizar correctamente el lenguaje tanto en la comunicación oral como escrita, y asimismo saber interpretarlo y comprenderlo en diferentes contextos. Debe permitir al alumno formarse juicios críticos, generar ideas y adoptar decisiones. En el caso de lenguas extranjeras, significa poder comunicarse en alguna de ellas de modo que se enriquezcan las relaciones sociales y favorezcan el desenvolvimiento en distintas situaciones.

#### Competencia matemática:

Implica poseer facilidad para utilizar y relacionar números, emplearlos en operaciones básicas y en razonamiento matemático para interpretar información, ampliar conocimientos y resolver problemas tanto de la vida cotidiana como del mundo laboral.

#### Competencia en el conocimiento y la interacción con el mundo físico:

Es la habilidad para desplegarse de forma autónoma en distintos ámbitos, como en la salud, el consumo o la ciencia, de modo que se sepa analizar, interpretar y obtener conclusiones personales en un contexto en el que los avances científicos y tecnológicos se hallan en continuo desarrollo.

#### Tratamiento de la información y competencia digital:

Esta competencia se refiere a la capacidad del alumno para buscar, obtener, procesar y comunicar información y transformarla en conocimiento. Esto supone destreza para acceder a la información y transmitirla en diferentes formatos, así como hacer uso de los recursos tecnológicos para resolver problemas reales de modo eficiente.

#### Competencia social y ciudadana:

Entre las habilidades de esta competencia se incluyen el conocerse y valorarse a uno mismo, saber comunicarse en diferentes contextos, expresar las ideas propias y escuchar las ajenas, comprendiendo los diferentes puntos de vista y apreciando tanto los intereses individuales como los del grupo, en definitiva aptitudes para participar activa y plenamente en la vida cívica.

#### Competencia cultural y artística:

Esta competencia se refiere a la capacidad de conocer, comprender, apreciar y considerar críticamente las distintas manifestaciones culturales o artísticas, así como saber emplear algunos recursos de la expresión artística para realizar creaciones propias.

Competencia para aprender a aprender:

Se refiere al aprendizaje a lo largo de la vida, es decir a la potencia para continuar aprendiendo de manera eficaz y autónoma una vez finalizada la etapa escolar. Esto implica, además de tener control de las propias capacidades y conocimientos y estar debidamente motivado, el saber utilizar adecuadamente estrategias y técnicas de estudio.

Autonomía e iniciativa personal:

Responsabilidad, perseverancia, autoestima, creatividad, autocrítica y control personal son algunas de las habilidades relacionadas con esta competencia, que permiten al estudiante emplear una visión estratégica de los retos y oportunidades a los que se debe enfrentarse a lo largo de su vida y le facilitan la toma de decisiones.

#### **4.20 La evaluación de las competencias, elementos previos**

El tema de la evaluación es uno de los que más ha provocado agitación en el ámbito docente. Durante décadas hemos acostumbrado “calificar” los trabajos de los estudiantes, y por lo tanto, emitir calificativos de sus acciones, actitudes e ideas. Sin embargo, debido a la cantidad de información que poseen los estudiantes actualmente, así como a los nuevos puntos de vista sobre la educación a nivel mundial, este paradigma tradicional ha entrado en crisis, en lo relacionado a la respuesta que la sociedad del conocimiento requiere de la educación básica, media y superior; pero no así en las prácticas docentes efectivas dentro un gran número de instituciones y, por consiguiente, dentro del aula. Atendiendo a esta situación actual que se vive en muchos países, ha surgido una propuesta desde el enfoque socio-formativo que mediante el trabajo conjunto está generando nuevas experiencias de evaluación en universidades, instituciones públicas y privadas, así como en organismos encargados de este importante proceso educativo.

La evaluación es un proceso inherente a la vida de las personas que buscan la mejora de su entorno y de sí mismas, y por ello es un tema que necesitamos abordar desde nuestra propia experiencia como docentes, padres de familia, empleados, empresarios, y cualquier otro rol que desempeñemos en la sociedad.

La evaluación de competencias por ser un proceso sistémico debe incluir instrumentos que permitan evidenciar el logro de las competencias bajo ciertos criterios establecidos puntualmente. Estos criterios o aprendizajes esperados deberán estar basados en las competencias genéricas o disciplinares previamente establecidas por la RIEMS de manera preferente, sin excluir la adaptación e innovación de otras conforme al contexto.

Ante esta situación, algunos de los docentes que han iniciado el proceso de cambio de paradigma en su actuar profesional se han encontrado con limitantes al realizar de forma sistemática esta evaluación, pues aunque podemos encontrar una serie de elementos teóricos de forma gratuita (internet) o mediante bibliografía, en ocasiones es complicado plasmar de forma didáctica este tema.

Cabe mencionar que el principal reto de la evaluación de competencias no son los instrumentos, sino los escenarios que permitan evidenciar el desempeño integral de las personas y los problemas del contexto planteados para su resolución. Por ello, uno de los obstáculos que dentro de nuestro rol docente nos corresponde enfrentar en muchas ocasiones, es la burocracia evaluativa, es decir, la elaboración de instrumentos que la mayoría de las ocasiones justifican el actuar docente más no dan elementos creíbles de la mejora de los estudiantes como puede observarse en la tabla comparativa entre la evaluación tradicional y la de competencias de este bloque. Es por esto, que te invitamos a dar un paso más en la implementación de las competencias dentro de tu aula: su valoración para la mejora.

#### **4.21 A modo de síntesis**

La evaluación constituye una parte consustancial de toda actividad humana y es en sí misma una actividad que se manifiesta como proceso. Es el proceso y resultado de juzgar la valía de un objeto o fenómeno de la realidad (sus características esenciales, sus manifestaciones

particulares, su devenir, su desarrollo actual y previsible), de acuerdo con criterios de referencia pertinentes a la naturaleza del propio objeto y a los propósitos que se persigan.

Como actividad, la evaluación posee una composición estructural y funcional: se realiza en función de objetivos o fines previstos, tiene un objeto: aquello que se evalúa, se realiza en interacción del evaluador con el objeto mediante un conjunto de acciones y operaciones que requieren el uso de procedimientos y medios adecuados, se organiza en ciertas formas, se efectúa en determinadas condiciones y tiene resultados que pueden coincidir en mayor o menor medida con los fines propuestos. Asimismo, supone fases de orientación, ejecución, control y regulación (de evaluación de la propia actividad).

Las concepciones sobre la evaluación han estado, con demasiada frecuencia, plagadas de reduccionismos que limitan y hasta distorsionan su práctica. La evaluación no es calificar (entendida como asignación de notas). La evaluación no es examinar (realizar exámenes). La evaluación trasciende a ambas acciones.

La calificación es una acción de la evaluación que permite expresar, de modo resumido, un juicio evaluativo y que se puede manifestar de maneras diversas (símbolos, adjetivos calificativos, informes descriptivos valorativos u otras). La nota juzga sumaria y concluyentemente un proceso continuo y no da cuenta de las fortalezas y las debilidades del aprendizaje del alumno, de su potencialidad para aprender, por lo que porta una información limitada. El examen es solo uno de los recursos o técnicas para obtener información sobre el aprendizaje de los estudiantes. En la práctica de la evaluación, tradicionalmente, se ha privilegiado en demasía la nota y el examen, de tal suerte que se aprende y se enseña en función del examen, para aprobarlo. Lo que interesa es la nota.

En síntesis, la evaluación educativa del aprendizaje debe orientarse a explorar, valorar y coadyuvar al desarrollo de las potencialidades de cada estudiante y a la búsqueda de vías de desarrollo a partir de la diversificación de oportunidades y espacios que la enseñanza y la propia sociedad brinda, o debe brindar.

#### **4.22 Epílogo: un nuevo marco conceptual para la medición y evaluación educativa**

Es evidente que en el campo de la medición y evaluación educativa se ha producido un cambio fundamental en su marco conceptual. La manera en la que en la actualidad se genera el discurso en esta área de conocimiento ha evolucionado profundamente incorporando un conjunto de nuevos elementos que apenas gozaban de consideración anteriormente, mientras que otros que habían ocupado un espacio nuclear de forma tradicional se han visto relegados a un papel más complementario.

Así y, en primer lugar, se ha observado un claro desplazamiento de los planteamientos evaluativos basados en exclusiva en los principios psicométricos a otros centrados en la evaluación educativa. Se ha pasado de la cultura de los tests a la cultura de la evaluación. Todo ello ha producido un abandono sistemático del término medición educativa a favor del más comprensivo de evaluación educativa.

Pero, sería injusto reducir el problema a un simple cambio tecnológico, el desplazamiento terminológico responde, por encima de todo, un conjunto de transformaciones de mucho mayor calado, de carácter paradigmático, así se han modificado nuestras concepciones sobre la naturaleza del aprendizaje, del rendimiento y de la evaluación. Nuestras definiciones sobre estos tres elementos son radicalmente distintas de las que se han ido sosteniendo de forma tradicional, también se ha incorporado el factor ético como un factor consustancial a la realidad evaluativa.

Nuestro mensaje a lo largo del artículo ha sido básicamente que la evaluación constituye una parte fundamental del hecho educativo y su uso se justifica en tanto en cuanto optimizamos su impacto sobre la calidad de los aprendizajes. Posiblemente y desde una postura más radical, podríamos decir que las actividades evaluativas no son sino actividades educativas y la distinción entre unas y otras es puramente metodológica o académica, pero en ningún caso establece diferenciación por razón de su naturaleza.

Pretendemos con esta aportación recapitular los elementos constitutivos de lo que podría constituir un nuevo marco para la medición y evaluación educativa. Es evidente que la mayoría

de ellos han ido apareciendo a lo largo de la última década, pero posiblemente es necesario significarlos de forma conjunta para establecer la línea argumental y estructural que los aglutina y poder orientar definitivamente nuestro pensamiento y acción hacia la necesidad de diseñar y establecer de forma generalizada un nuevo paradigma para la evaluación educativa.

La pretensión clásica de la medición ha consistido históricamente en tratar de cuantificar los atributos que están presentes en objetos o individuos. Con ello se intenta objetivar nuestros juicios valorativos respecto de los mismos y facilitar su estudio y su conocimiento.

Con la medición se intenta hacer comparables los individuos en relación a aquellas propiedades que previamente hemos analizado. La medición empezó a aplicarse a las Humanidades a partir del siglo XIX cuando se intentó introducir este tipo de conocimientos en la misma metáfora en la que ya estaban situadas las denominadas disciplinas científicas. Medir era la condición previa necesaria para poder aplicar el método científico a cualquier área específica de conocimiento que pretendiera conseguir estatus científico.

Con todo, la analogía aplicada a las Humanidades, entendiendo que las características medibles de estos conocimientos eran de naturaleza similar a otros saberes como, pongamos por caso la Física, esta hoy en clara crisis y se han ido evidenciando un gran número de limitaciones, algunas de carácter conceptual, otras instrumental y finalmente también de tipo metodológico, que trataremos de presentar brevemente a continuación.

#### Limitaciones de carácter conceptual

La medición educativa adoptó como propios en sus inicios los principios de la psicometría, sin embargo hemos de considerar que la psicometría como construcción científica se realizó a partir de los trabajos en torno a la inteligencia y a su medición.

La noción que subyacía era que la inteligencia era una característica de naturaleza similar a otras características físicas de las personas y que podía ser también observada y consecuentemente medida.

Su propia naturaleza establece su primera dificultad para trasladarse sin más al entorno educativo. Así desde la psicometría, se considera que los atributos medidos son de carácter

fijo, estableciéndose un concepto de la inteligencia muy cristalizado y poco fluido, en contraste con el uso que la medición debería adoptar en el contexto educativo, donde la medición de los aprendizajes debería estar orientada a su propio desarrollo.

La medición de cualquier aprendizaje incide inevitablemente en su proceso de cambio y mejora. La medición adopta por tanto un carácter mucho más dinámico cuando se aplica a la educación.

Otro factor de la psicometría que puede considerarse como una limitación desde la perspectiva educativa, es la necesaria interpretación de las puntuaciones en función de la norma establecida por el propio grupo. Así en el marco de la psicometría clásica las puntuaciones de cualquier sujeto se establecen por comparación con sus pares y por tanto la ejecución del sujeto en la prueba siempre se expresa en términos relativos, renunciando por las características de su propio diseño a establecer ningún tipo de interpretación absoluta del atributo medido.

#### Limitaciones de carácter instrumental

El enfoque clásico de la medición ha basado casi en exclusiva la recogida de la información en la aplicación de tests estandarizados de carácter presuntamente objetivo. Modelo de actuación que incluso en la actualidad sigue siendo el preponderante.

Esta perspectiva tal como se plantea, nos proporciona un tipo de información cuantificada que es fundamentalmente útil para comparar las realizaciones de unos grupos de estudiantes frente a otros y no facilita, en realidad, otros tipos de información también relevante que nos permitan una gestión más eficaz de los juicios extraídos sobre el conocimiento de nuestros alumnos.

#### Limitaciones de carácter metodológico

Los dos pilares metodológicos sobre los que se sustenta la calidad de los procesos de medición son la validez y la fiabilidad. Vamos a continuación a analizarlos desde una perspectiva crítica.

#### En torno a la validez

Le corresponde a la fiabilidad la categoría de ser el término más controvertido y sin embargo la validez ha sido, de entre los que conforman la medición y evaluación educativa, el término más ampliamente tratado a lo largo de los últimos años y posiblemente Messick (1994, 2004) ha sido el autor que más profusamente ha reflexionado sobre el mismo. En líneas generales,



desde su perspectiva, la validez se debe fundamentar sobre dos tipos de aportaciones: en las evidencias que aporta el enfoque de validez de constructo y en las consecuencias sociales que se derivan del uso de las pruebas.

En torno a la fiabilidad

Si bien la reconceptualización de la validez la hemos resuelto convirtiendo el tema en una cuestión de priorización y especificación de las responsabilidades que competen al diseñador, al administrador, al editor y al usuario de la prueba, repensar la fiabilidad precisa de una aproximación mucho más radical.

La evaluación educativa no opera bajo la asunción de unidimensionalidad de los constructos que pretende medir, la distribución de las puntuaciones correspondiente a los resultados difícilmente se ajusta a la curva normal, usualmente trabaja con muy pocos elementos y las condiciones de evaluación no son fácilmente estandarizables. Con todo ello pretendemos señalar que la aproximación estadística, usualmente aplicada en la estimación de la fiabilidad de los tests estandarizados es simplemente inadecuada para las pruebas basadas en ejecuciones.

#### **4.23 Cambio de paradigma en la medición y evaluación educativa**

##### **BASES PARA UN CAMBIO DE PARADIGMA**

La Psicometría clásica original se basó en la teoría de la inteligencia para fundamentarse, ante su insuficiencia en el campo educativo se buscó posteriormente solución en los modelos conductistas y aparecieron todo el universo de las pruebas denominadas objetivas, que tampoco solucionaron la mayor. El desarrollo de un nuevo paradigma debería huir claramente de cimentarse en alguna teoría en concreto y más bien perseguir fundamentalmente ayudarnos a comprender los nuevos modelos y procesos de aprendizaje.

Desde el punto de vista histórico Wood (1986) cita el trabajo de Glaser en 1963 sobre las pruebas referidas al criterio, como el momento clave de inicio de la separación respecto de los planteamientos métricos clásicos y el tránsito conceptual de pasar de la medición a la evaluación educativa como concepto nuevo e integrador. Para Wood se inicia en ese momento la construcción de un nuevo paradigma para la medición y evaluación educativa.

En el marco del paradigma emergente se buscaba diseñar las pruebas de forma que la información obtenida permitiera analizar el sujeto por referencia a sí mismo y no por referencia a los demás individuos del grupo y facilitar el uso de los resultados de forma constructiva, identificando las fortalezas y debilidades de los individuos de manera que orientara la acción para su mejora educativa.

Según Wood (1986) el nuevo paradigma se podría definir por su acuerdo con los siguientes principios:

- Trata sobre el rendimiento del sujeto en relación consigo mismo, más que por referencia al rendimiento de otros.
- Busca comprobar la competencia más que la inteligencia.
- Tiene lugar bajo condiciones relativamente incontroladas y consecuentemente no produce datos conductualmente correctos.
- Busca la «mejor ejecución» más que la «típica ejecución»
- Es más efectivo cuando las normas y regulaciones características de los tests estandarizados son más flexibles.
- Permite una mirada más constructiva respecto de la evaluación donde el objetivo es ayudar al sujeto más que sentenciarlo.

En los últimos años los cambios en el paradigma se han hecho evidentes en muchos aspectos, a lo largo de nuestro artículo trataremos de señalarlos, sin embargo vamos a indicar uno inicial que entiendo corresponde citar en este momento.

Nos referimos al cambio muy visible que se ha producido en el paulatino cambio de denominación del área de conocimiento. Hemos pasado de la medición educativa a la evaluación educativa (entendida en términos de «assessment», recordemos que la evaluación tiene otras acepciones en las que no entramos en esta reflexión). Los procesos que antes estaban vinculados con la medición se han ido modificando y traspasado al paradigma de la evaluación, de tal forma que de manera generalizada se usa casi en exclusiva el término evaluación como único e inclusivo.

Gipps (2000), justifica el cambio a partir que mientras la medición implicaba «cuantificación precisa», la evaluación actual no está totalmente interesada en este tipo de proceso y ha encontrado nuevas vías de capturar la información evaluativa sustantiva y de garantizar su calidad.

#### **4.24 Aceptaciones en torno al concepto de medición y evaluación educativa. En busca de una definición comprensiva del termino**

En todo proceso educativo se hace estrictamente necesaria la presencia de la evaluación. Y es que esta herramienta constituye un abanico abierto tanto para los alumnos como para los profesores, llevando a un desarrollo de las fortalezas y a cambio en las debilidades. Cuando buscamos definirla nos encontramos con un listado bastante amplio de información y autores. He aquí, que buscando justamente una definición para la evaluación educativa es que llegamos a un punto de coincidencia.

Podemos decir que la evaluación educativa es una herramienta de gran importancia para el proceso de aprendizaje. Debido a que la educación es un proceso largo y que la evaluación como herramienta forma parte de él, nace la necesidad de ser no solo sistemática, sino también permanente y objetiva en la información que recaba. Toda esa información que se pueda obtener a través de esta herramienta ayuda a orientar al alumno a partir de la exanimación de los resultados con respecto a los objetivos previamente planteados. Además de lo anterior entrega evidencia de los cambios en los alumnos y del grado que han logrado en éste. A partir de la valoración que el docente pueda dar a la realidad de sus alumnos, después de haber reunido información útil y descriptiva, es tiempo centrarse en un proceso de identificación, remodelación y tratamiento de datos que construirán el camino para llegar al conjunto de actividades que conducen a un juicio en función de criterios. Pero la genialidad de la evaluación va aun más allá debido a que permite criticar y revisar los planes, programas y métodos con el fin de mejorar el proceso educativo.

Cuando nos referimos al quehacer educativo, necesario distinguir entre programación y evaluación. Y es que no basta llevar a cabo la programación sino que es esencial evidenciar claramente hasta donde se han cumplido aquellos objetivos definidos con anterioridad y sobre

lo mismo tomar las decisiones más adecuadas. Es así que cuando evaluamos estamos haciendo un contraste entre lo que se mide y lo que se quiere lograr.

El criterio es un elemento que sin duda debe ir de la mano de toda evaluación. Para definirlo con más facilidad, debemos decir que el criterio es un objetivo circunscrito previamente de acuerdo a lo que se espera de los alumnos.

El criterio necesita los siguientes componentes:

Rendimiento del alumno en función a sus posibilidades.

Progreso (relación entre el rendimiento actual y rendimiento anterior).

Norma, límite o meta exigida (entendida como el mínimo que se debe exigir al alumno).

#### **4.25 En torno a la validez**

La validez de una prueba indica el grado en que un instrumento mide aquello que pretende medir y proporciona información adecuada en relación a su uso y utilización. Cuando nos planteamos la validez de una prueba, hemos de responder a las siguientes preguntas:

- ¿Mide el instrumento aquello que queremos medir?
- ¿En qué grado lo mide?
- ¿Proporciona la información necesaria para nuestros propósitos evaluativos?

La validez no es una propiedad general del instrumento de medida sino que depende del propósito que se persiga con su utilización, de manera que no se puede hablar de una validez general sino relativa a los objetivos de aprendizaje que queremos medir y al tipo de decisión que se quiera tomar con la información que nos proporcione el instrumento.

#### **4.26 En torno a la fiabilidad**

Este concepto hace referencia a la exactitud y precisión de resultados obtenidos con la aplicación de una prueba determinada. El objetivo es reducir el margen de error en la medida de lo posible. La fiabilidad depende en buena parte del número de datos (o preguntas). Como

no se puede preguntar todo a todos, habrá que tener más datos de aquello que se juzgue más importante (y esto supone clarificar qué es más y menos importante).

#### **4.27 En torno a la naturaleza de los aprendizajes**

A lo largo de los últimos años, el aprendizaje ha pasado cada vez más a ocupar un lugar central por una serie de razones importantes que encuentran eco en la política, así como en la educación, en muchos países, tal como indican Dumont e Istance (capítulo I). Dichas razones definen los objetivos de este importante volumen desde la labor sobre los Ambientes Innovadores de Aprendizaje elaborada por el Centro para la Investigación y la Innovación Educativas (CERI).

Diferentes factores a nivel mundial ponen de manifiesto cada vez más lo que algunos llaman las “competencias del siglo XXI”. La cantidad y la calidad del aprendizaje se han vuelto centrales y los enfoques educativos tradicionales resultan insuficientes.

Factores similares ayudan a explicar la gran atención que se ha prestado a la medición de los resultados del aprendizaje (incluido el trabajo desarrollado por el Laboratorio Latinoamericano de Evaluación de la Calidad [LLECE] y el Programa Internacional para la Evaluación del Estudiante [PISA, por sus siglas en inglés]) en los dos últimos decenios, y que a su vez genera aún más atención en el aprendizaje. Para ir más allá del diagnóstico de los niveles de rendimiento y las deficiencias, y avanzar hacia un cambio conveniente, es necesario comprender mejor la forma en que las personas aprenden de la manera más eficaz posible.

El rápido desarrollo y la omnipresencia de la tecnología de la información y la comunicación (TIC) amplían los límites de las posibilidades educativas. Sin embargo, las grandes inversiones realizadas en recursos digitales no han revolucionado los ambientes de aprendizaje. Comprender la manera en que podrían hacerlo exige que se preste atención a la naturaleza del aprendizaje.

La sensación de alcanzar los límites de la reforma educativa invita a realizar un nuevo hincapié en el aprendizaje en sí: la educación se ha reformado una y otra vez en la mayoría de los países de la región, lo que lleva a muchos a plantearse la necesidad de nuevas formas de influir en la interfaz del aprendizaje y la enseñanza.

La base de investigación sobre el aprendizaje se ha ampliado notablemente, pero numerosos investigadores observan lo inadecuada que es la manera en que suelen aplicarse las conclusiones de las ciencias del aprendizaje en las escuelas. Al mismo tiempo, buena parte de la investigación sobre el aprendizaje está desconectada de la realidad de la práctica educativa, así como de la formulación de políticas.

A lo largo de los últimos años, el aprendizaje ha pasado cada vez más a ocupar un lugar central desde el punto de vista político por una serie de razones importantes. Este volumen, que es tanto una revisión de investigaciones como un análisis de las implicaciones de la investigación científica en la planificación educativa, está estrechamente definido por estos cambios. En el presente capítulo se brindan detalles sobre los desarrollos contemporáneos que sientan las bases para los siguientes capítulos. Tales desarrollos exigen aprovechar los conocimientos sobre el aprendizaje y aplicarlo de manera más sistemática a la educación. El capítulo explica por qué esos desarrollos presentan argumentos en favor de una atención especial al nivel micro de los ambientes de aprendizaje y por qué esto es algo que debe avanzar hacia el futuro con un fuerte foco en la innovación.

#### **4.28 En torno a la lógica de la evaluación**

Las evaluaciones se siguen haciendo, lamentablemente, al final de los procesos de enseñanza. Se interrumpe el aprendizaje para comprobar si se ha aprendido. La evaluación no avanza en paralelo a los aprendizajes, se realiza al final, como si fuera su coronación, su colofón, su cierre. Y no su natural correlato.

Y luego vienen los informes, las calificaciones (que algunos confunden con la propia evaluación). Los alumnos y las alumnas llevan las calificaciones a las familias. Lo que les preguntan los padres a sus hijos e hijas no es, probablemente, si lo que han aprendido les ha hecho mejores personas o si han ayudado a los demás a aprender o si han respetado a quienes les han enseñado o si han agradecido sus esfuerzos a quienes les han ayudado en el empeño. Les preguntan por los resultados. Exclusivamente. Porque eso es lo que de verdad importa.

Ahora bien, para alcanzar el éxito, hay que satisfacer las demandas de los evaluadores. Hay que responder a lo que se pregunta y hay que hacerlo de la forma pertinente para que la respuesta encaje en las expectativas de quien ha de decir si está bien o está mal. Por eso la evaluación encierra mucho poder.

Mientras más se asciende en la jerarquía académica (desde el alumnado al ministerio) va disminuyendo la evaluación y aumentando el sueldo. Mientras más se desciende, aumenta la evaluación y baja la remuneración o, incluso, hay que pagar. La evaluación, como decía, contiene poder. Y el poder tiene sus leyes.

¿Cómo se consigue el éxito en los denostados exámenes? En muchos casos, repitiendo lo que se ha leído o lo que se ha escuchado. Pero sin tener en cuenta cómo ha sido de clara, rica y motivadora esa enseñanza. No sé dónde he leído que si los profesores enseñasen geografía a las palomas mensajeras, muchos de los mensajes no llegarían a su destino. Con perdón de los profesores y profesoras de geografía, siempre competentes y esforzados, salvo las evitables excepciones.

He conocido excelentes profesionales que han enseñado geografía con pasión y paciencia. Uno de ellos me contó hace años una simpática anécdota sobre evaluación de la asignatura. Delante de un mapa físico de España colgado en la pared del aula, le pidió a un niño que indicase dónde nacía el río Nalón y dónde desembocaba.

El niño, muy desenvuelto, indicó con un puntero que nacía en la ría de Pravia (Mar Cantábrico, en la parte superior del mapa) y que desembocaba en la Fuente Nalona (Puerto de Tarna, unos centímetros más abajo). Lo había estudiado con ahínco.

– Se nota que has estudiado, dijo el profesor, pero no has entendido bien las dos referencias. Es exactamente al revés, dijo el profesor: nace en la Fuente Nalona y desemboca en el Mar Mediterráneo.

El niño, con mucho aplomo y cierto enfado, como indicando algo obvio, dijo:

– ¡Sí, hombre, el agua va a ir hacia arriba!

La evaluación nos ofrece testimonios como el de este pequeño geógrafo. Aplican su peculiar lógica. Porque no digan lo que nosotros esperamos no quiere decir que no razonen.

Pondré otro ejemplo que me brindó la madre de un alumno al terminar una conferencia sobre evaluación en la ciudad Mar del Plata de Argentina. Me contó que a un hijo suyo le había pedido la profesora de geografía que dibujase durante el fin de semana un accidente costero. El chico dibujó una playa en la que se podía ver un adulto con un niño al lado. Cuando entregó el trabajo, la maestra le preguntó:

- ¿Dónde está el accidente costero?
- Ahí está, profesora.
- No lo veo por ninguna parte.
- Está ahí, en el centro, en esas dos figuras. Un día fui a la playa con mi papá y me rompí una pierna.

Podría poner ejemplos de otras asignaturas. Múltiples ejemplos en los que los alumnos siguen el rastro de la lógica. De su lógica particular.

Me preocupa mucho que la evaluación de los aprendizajes tenga un fuerte componente memorístico, repetitivo, mecánico. Sin verdadera comprensión. No estoy contra del cultivo de la memoria ya que es la sustancia de la vida humana. No podríamos vivir sin memoria. Lo que cuestiono es la repetición mecánica de los contenidos. Lo que se ha dado en llamar el aprendizaje memorístico.

Un profesor de historia examinaba de esta forma, de modo que todos sus alumnos sabían que para tener una buena nota debían repetir literalmente lo que el texto decía. Si (es un caso real) el libro decía “Los vándalos entraron en la península en oleadas sucesivas”, la pregunta sobre cómo entraron los vándalos en la península, solo podía ser correctamente contestada con la frase literal del libro: “en oleadas sucesivas”. Si la contestación era fidedigna pero no literal (“a caballo”, “en momentos diferentes”, “en tropel”, “arrasando”, “paulatinamente”...) era considerada un error.



En un examen oral de Historia le preguntaron a un alumno:

- ¿Por qué fueron expulsados los judíos de la Península?
- Porque no querían dejarse hacer fotos, contestó.
- No es posible. Si ni siquiera había cámaras, dijo el profesor.
- El libro que yo tengo lo dice claramente.
- Trae el libro y lee despacio delante de mí.
- El chico leyó fijándose ahora en una “c” que antes había pasado inadvertida.
- Los judíos fueron expulsados de la Península porque no querían retractarse.

Aquella maestra dejó plasmado el error del aprendizaje memorístico en aquella advertencia que formuló a sus alumnos y alumnas: “Atención, niños. Esto que voy a explicar es muy importante. Tenéis que repetirlo literalmente en el examen. Bueno, si alguien no es capaz de decirlo literalmente, lo puede expresar con sus propias palabras”.

No. No. No. Será al revés. Si alguien no lo puede decir con sus propias palabras (que es más difícil y deseable) que lo repita literalmente.

Creo que es una trampa la excusa que algunos utilizan para hacer una evaluación de esta naturaleza cuando dicen que se trata de preparar a los escolares para realizar en etapas superiores ese tipo de evaluación. En lugar de preparar para la guerra, ¿no sería más lógico y más justo acabar con la guerra?

¿Es educativa la evaluación? Esa es, a mi juicio, la cuestión central. Si es educativa no es solo porque se centra en dimensiones diversas de la educación sino porque educa a quien la hace y a quien la recibe. ¿Cuándo educa la evaluación? Cuando respeta, ayuda, motiva, hace pensar, facilita el aprendizaje, mejora la enseñanza y despierta el deseo de saber.

#### **4.29 Criterios de calidad alternativos para las pruebas de ejecución**

##### **a. Lista de verificación de desempeño**

Es una relación descriptiva que enumera, paso a paso, las actividades de una tarea respetando su secuencia. Se usa para evaluar procesos y cotejar entre “sí” o “no”; o entre “aceptable” o “inaceptable”; o algunas otras designaciones similares.

##### **b. Escalada Estimativa de Desempeño**

Es una lista de actividades o aspectos a evaluar llamados variables. Estas variables se clasifican juzgando su nivel de efectividad, según una graduación establecida. Es apropiada para evaluar la calidad de un trabajo, particularmente cuando se evalúa un producto terminado.

#### **PRUEBAS DE EJECUCIÓN – MODALIDAD PRODUCTO TERMINADO**

Los instrumentos a usar pueden ser listas de verificación y/o escala estimativa.

##### **a. Lista de verificación de producto terminado**

Es uno de los instrumentos más usados para evaluar objetivos del dominio psicomotriz. La lista de verificación para producto terminado no es una relación de acciones, sino un listado de las características que se espera encontrar en lo que el aprendiz haya elaborado.

##### **b. Escala estimativa de producto terminado**

Evalúa el aprendizaje de un participante a través del análisis de un producto de su trabajo. Especialmente, cuando no sólo nos interesa saber si el producto presenta ciertas características, sino además, en qué grado las presenta. Son instrumentos que usa el formador, para calificar después de terminada la ejecución del alumno.

#### **4.30 El factor ético**

En toda práctica profesional podemos encontrar implicaciones éticas de un mayor o menor grado, en este sentido es lógico pensar que las actividades que tengan que ver con prácticas sociales, con contacto humano y educativo pueden tener implicaciones éticas a las que hay que darles una mención especial.

En el ámbito educativo si no se llevase a cabo una evaluación, a través de una sana vigilancia moral, se podría caer en la injusticia, abuso de poder... Por ello, es necesario velar para que los valores y principios morales se vean latentes en la realidad educativa.

La actuación de los profesores, en la que vamos a incluir la evaluación académica, tiene una inevitable influencia ideológica y moral, es cierto que multitud de profesores afirman mantener una cierta imparcialidad a la hora de desarrollar su labor, pero es indiscutible que hay muchos aspectos que de una forma inconsciente o simplemente por puro desconocimiento escapan de sus manos. Ésta es la causa, por lo cual es necesario tener en cuenta, muchos de estos aspectos como es el caso del etiquetaje, intereses personales, ideas preconcebidas, influencia de rumores, etc...e intentar actuar en consonancia ética con nuestra labor profesional. Sobre todo, siendo conscientes de que cualquier interferencia o contaminación causada por algún aspecto moral puede dar como resultado una conclusión evaluadora errónea influyendo de este modo todo el proceso evaluativo.

En relación con los temas éticos que trata la evaluación educativa, como en toda puesta en práctica, se pueden percibir tanto luces como sombras. Esto puede deberse a la focalización que ocurre cuando se lleva a cabo la evaluación en el día a día. El objetivo de señalar estas debilidades, no es por buscar detractores de ética en la evaluación educativa, sino para todo lo contrario, es decir, para crear conciencia y dar a conocer dónde hay que prestar más atención.

Las cuestiones éticas y morales que afectan directamente a las personas que han sido sometidas a la evaluación son absolutamente imprescindibles. El motivo de esta dejadez en estos aspectos tan relevantes puede deberse al poder y la autoridad que demuestran los propios evaluadores en relación con el estado, la escuela, los docentes y aquellos dados a la evaluación por los distintos agentes externos como puede ser el caso de los asesores y evaluadores.

El estudio de la reflexión de la puesta en práctica, implica algo más que el mero cálculo de los medios para alcanzar un fin. El trabajo educativo participa de manera activa en la vida de las personas siendo un factor influyente en el desarrollo de las personas. Por este motivo, no existe unanimidad en las prácticas didácticas ya que cada una pretende promover unos valores u otros dependiendo de la realidad escolar

Tal y como hace referencia Meirieu (2001) existen prácticas didácticas, entendiéndose como modos de funcionar de las instituciones educativas, de las que uno sólo puede apartarse prescindiendo de los demás e incluso, algunas veces, pisándolos. Es por este motivo, por lo que la idoneidad ética de los medios al igual que ocurre con la técnica educativa, se transforma en aspectos relevantes. Por ello, es vital hacer una unión entre el profesionalismo con la comprensión, con el marco de identidad de la responsabilidad adquirida junto con la obligación de cumplir el deber profesional sin perder de vista el compromiso social.

Otras omisiones se dan en la objetividad de los profesionales. Estas carencias se pueden producir por la falta de competencia profesional o de manera consciente e intencionada. En la primera causa, se cometen errores en el diseño incorrecto de evaluaciones, es decir, en el instrumento de evaluación o inexistente relación entre los contenidos impartidos y los contenidos evaluados. Esta debilidad hace que el profesional presente un déficit en su formación, aunque éste no sea consciente de ello, no le exime de su responsabilidad. En relación con el segundo motivo, vinculado con la intencionalidad se produce cuando el profesional decide beneficiar a un evaluado sobre otro, sin razones sólidas viéndose afecta la evaluación por la subjetividad de los intereses propios del evaluador. El error se centra en la ponderación de los resultados por voluntad o comisión que se produce cuando un profesional incurre en una falta o culpa con el fin de favorecer a una persona.

Las debilidades éticas, teniendo en cuenta la falta de atención que se ha tenido en los aspectos relacionados con los valores esenciales para la calidad tanto del docente como del estatus de su grupo profesional, deben sentir la obligación de brindar sus servicios a personas o grupos, caracterizándose de poseer un comportamiento digno, responsable, distinguido, trascendente y comprometido.

## **BIBLIOGRAFIA**

### **BASICA**

[https://www.uv.mx/personal/jomartinez/files/2011/08/LA\\_EVALUACION\\_EDUCATIVA.pdf](https://www.uv.mx/personal/jomartinez/files/2011/08/LA_EVALUACION_EDUCATIVA.pdf)

[http://online.aliat.edu.mx/adistancia/Calidad/unidad6/lecturas/TEXTI\\_S6\\_AnaDatos.pdf](http://online.aliat.edu.mx/adistancia/Calidad/unidad6/lecturas/TEXTI_S6_AnaDatos.pdf)

[http://www.cucs.udg.mx/revistas/edu\\_desarrollo/anteriores/8/008\\_Vincenzi.pdf](http://www.cucs.udg.mx/revistas/edu_desarrollo/anteriores/8/008_Vincenzi.pdf)

[https://www.uv.mx/personal/jomartinez/files/2011/08/LA\\_EVALUACION\\_EDUCATIVA.pdf](https://www.uv.mx/personal/jomartinez/files/2011/08/LA_EVALUACION_EDUCATIVA.pdf)

<http://www.cise.espol.edu.ec/sites/cise.espol.edu.ec/files/pagina-basica/Nota%20te%CC%81cnica%20n.%C2%BA%209%20-%20%20C2%BFCua%CC%81les%20son%20los%20tipos%20de%20evaluacio%CC%81n%20educativa%3F.pdf>

<http://www.cise.espol.edu.ec/sites/cise.espol.edu.ec/files/pagina-basica/Nota%20te%CC%81cnica%20n.%C2%BA%209%20-%20%20C2%BFCua%CC%81les%20son%20los%20tipos%20de%20evaluacio%CC%81n%20educativa%3F.pdf>

[https://books.google.com.mx/books?id=KCG6fO5jH9kC&pg=PA311&lpg=PA311&dq=medicion+y+evaluacion+educativa+joan+mateo+pdf&source=bl&ots=XdV32phyhE&sig=ACfU3U0FyOgrj6RZjX6mGtk4gPwcDwoejw&hl=es-419&sa=X&ved=2ahUKEwiw65yW2e\\_sAhUEiqwKHbnMDDA4FBD0ATACegQIARAC#v=onepage&q=medicion%20y%20evaluacion%20educativa%20joan%20mateo%20pdf&f=false](https://books.google.com.mx/books?id=KCG6fO5jH9kC&pg=PA311&lpg=PA311&dq=medicion+y+evaluacion+educativa+joan+mateo+pdf&source=bl&ots=XdV32phyhE&sig=ACfU3U0FyOgrj6RZjX6mGtk4gPwcDwoejw&hl=es-419&sa=X&ved=2ahUKEwiw65yW2e_sAhUEiqwKHbnMDDA4FBD0ATACegQIARAC#v=onepage&q=medicion%20y%20evaluacion%20educativa%20joan%20mateo%20pdf&f=false)

[https://books.google.com.mx/books?id=KCG6fO5jH9kC&pg=PA311&lpg=PA311&dq=medicion+y+evaluacion+educativa+joan+mateo+pdf&source=bl&ots=XdV32phyhE&sig=ACfU3U0FyOgrj6RZjX6mGtk4gPwcDwoejw&hl=es-419&sa=X&ved=2ahUKEwiw65yW2e\\_sAhUEiqwKHbnMDDA4FBD0ATACegQIARAC#v=onepage&q=medicion%20y%20evaluacion%20educativa%20joan%20mateo%20pdf&f=false](https://books.google.com.mx/books?id=KCG6fO5jH9kC&pg=PA311&lpg=PA311&dq=medicion+y+evaluacion+educativa+joan+mateo+pdf&source=bl&ots=XdV32phyhE&sig=ACfU3U0FyOgrj6RZjX6mGtk4gPwcDwoejw&hl=es-419&sa=X&ved=2ahUKEwiw65yW2e_sAhUEiqwKHbnMDDA4FBD0ATACegQIARAC#v=onepage&q=medicion%20y%20evaluacion%20educativa%20joan%20mateo%20pdf&f=false)

<https://sites.google.com/site/tecninvestigacionsocial/temas-y-contenidos/tema-4-las-tecnicas-estructurales-entrevista-grupo-de-discusion-observacion-y-biografia/tecnicas-de-observacion/observacion-sistemica>

### **COMPLEMENTARIA**

[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1665-109X2010000200005](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-109X2010000200005)

<https://www.uls.edu.sv/sitioweb/component/k2/item/311-la-metaevaluacion-mas-alla-de-lo-visible-por-que-evaluar-las-evaluaciones>

[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-2698201000100009](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-2698201000100009)

<https://blog.ecagrupoeducativo.mx/eca/las-8-competencias-educativas-mas-importantes-para-los-alumnos>