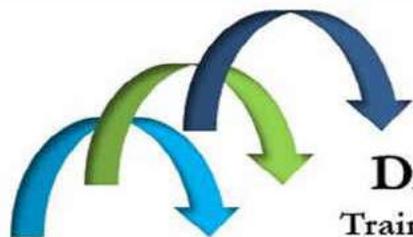


# Instrumentos y Técnicas de Investigación Educativa

Un Enfoque  
Cuantitativo y Cualitativo  
para la  
Recolección y Análisis  
de Datos



**DANAGA**  
Training and Consulting

Quality, education and continuous improvement

---

Carlos Ruiz Bolívar

# Instrumentos y Tecnicas de Investigación Educativa

Un Enfoque Cuantitativo y Cualitativo para la  
Recolección y Análisis de Datos

Carlos Ruiz Bolívar

Tercera Edición, Octubre 2013

© Carlos J. Ruiz Bolívar

© Copyright by DANAGA Training and Consulting  
Houston, Texas • USA

Reservado todos los derechos.

Este libro no puede ser reproducido total o parcialmente, por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética o cualquier almacenamiento de información y sistema de recuperación sin permiso escrito del autor.



# *Dedicatoria*

*A Bettys Rosario, Carlos Eduardo y Anangela*

# Contenido

DEDICATORIA

AGRADECIMIENTOS

PRÓLOGO A LA TERCERA EDICIÓN

INTRODUCCIÓN

PARTE I INSTRUMENTOS DE INVESTIGACION CUANTITATIVA

1. Instrumentos de Medición
2. Confiabilidad y Validez
3. Construcción de Escalas de Actitudes
4. Pruebas de Rendimiento Académico
5. Pruebas Formales e Informales de Rendimiento Académico
6. Construcción de Cuestionarios
7. Análisis Factorial

PARTE II TÉCNICAS DE ANÁLISIS CUALITATIVO

8. Observación Participante y Entrevista en Profundidad
9. Grupo Focal y Observación Flotante
10. Análisis de Datos Cualitativos

REFERENCIAS

# Agradecimientos

Escribir un libro en el contexto universitario es una tarea compleja en la cual participan, directa e indirectamente, muchas personas (estudiantes, colegas, amigos, familiares y editores, entre otros). Esta obra no ha sido la excepción. Estamos en deuda de gratitud permanente con nuestros estudiantes quienes sin estar conscientes de ello, han constituido la fuente de inspiración primaria que nos ha llevado a dedicar esfuerzos y tiempo extras, más allá de la rutina de la cátedra, para concluir este trabajo. Para ellos ha sido escrita, en primer lugar, esta obra.

A mis colegas de la UPEL y de otras instituciones de educación superior, quienes me han animado a continuar con este esfuerzo. A los colegas que tuvieron la gentileza de leer las versiones preliminares de esta obra y de hacer sugerencias valiosas para su mejoramiento. Mención especial, por su contribución, merecen los profesores: Fanny Torres de Giménez (+), Bettys Arenas de Ruiz y Luis Gómez Vásquez, quienes leyeron varios de los capítulos preliminares de esta obra. No obstante, como debe suponer el lector, los conceptos emitidos en este trabajo son de nuestra absoluta y exclusiva responsabilidad.

# Prólogo a la Tercera Edición

Nos sentimos muy satisfechos con la respuesta que los usuarios han dado a las dos ediciones anteriores de esta obra, cuyo propósito inicial era servir de ayuda a estudiantes y docentes que se inician en el campo de la investigación en ciencias sociales (educación, psicología, sociología, ciencias políticas) en cuanto al uso de los procedimientos adecuados para la construcción de instrumentos de investigación psicométricamente válidos. Con motivo de esta tercera edición hemos aprovechado la oportunidad para hacerle una revisión en profundidad y ampliar su alcance. En el primer caso, hemos hecho algunas correcciones de forma y fondo en el texto, se han mejorado algunos ejercicios y se han revisado todos los capítulos; mientras que en el segundo hemos atendido a las sugerencias de algunos usuarios quienes nos han planteado la conveniencia de que el libro incorpore, además, las técnicas de recolección y análisis de información y datos cualitativos.

Atendiendo a esa solicitud, hemos organizado la obra en dos partes principales, la primera referida al diseño y desarrollo de los instrumentos para la investigación educativa bajo el paradigma positivista; y la segunda, relacionada con el uso de las técnicas de recolección y análisis de información y datos en el contexto del paradigma naturalista-interpretativo. Con este nuevo alcance, esperamos ofrecer a los usuarios una obra actualizada que responda a las necesidades y expectativas de un mayor número de usuarios. Estas dos partes que conforman la estructura de la obra están precedidas por una Introducción referida al tema de los paradigmas,

enfoques y métodos de investigación que sirve de marco referencial para el mejor diseño y/o selección de los instrumentos y técnicas para la recolección y análisis de datos bajo los enfoques cuantitativo y cualitativo, según sea el caso.

Por otra parte, la obra mantiene el mismo estilo de siempre: sencillo, psico-pedagógicamente diseñado y con una orientación didáctica que le permite al usuario usarla como una guía metodológica a los fines de resolver las dudas relacionadas con los enfoques y procedimientos para la recolección y análisis de datos en la investigación social y educativa.

La primera parte de la obra comprende siete capítulos que corresponden a los aspectos teórico-prácticos del proceso de construcción de instrumentos en el contexto de la investigación en ciencias sociales y educación. El capítulo 1 está referido a los aspectos siguientes: instrumentos de medición, concepto de medición, medición y realidad, niveles de medición, clasificación de los instrumentos de medición y fases para la construcción de un instrumento de medición.

En el Capítulo 2, se enfatizan aspectos tales como: concepto e importancia de la confiabilidad, el modelo clásico de confiabilidad, tipos de confiabilidad, interpretación del coeficiente de confiabilidad y mejoramiento de la confiabilidad de una prueba. La segunda parte del capítulo se refiere al tema de la validez e incluye aspectos tales como: concepto e importancia de la validez, tipos de validez, factores que afectan la validez de un instrumento y relación entre validez y confiabilidad.

El Capítulo 3, toma en cuenta los aspectos siguientes: el estudio de las actitudes como constructo psicológico, conceptualización y medición de las actitudes, métodos de construcción de escalas, el método de rangos sumados

y desarrollo de la escala de actitud hacia la innovación educativa (ESCAHINE).

El capítulo 4 trata el tema de las Pruebas sobre el Rendimiento Académico (I) e incluye los aspectos siguientes: las pruebas escolares a través del tiempo y su relación con la evaluación educacional, definición, características técnicas y clasificación de las pruebas de rendimiento académico, interpretación de los resultados de las pruebas de aprovechamiento escolar. En el capítulo 5 también se trata el tema de las Pruebas sobre el Rendimiento Académico (II), pero referido a los tipos de pruebas: hechas por el docente, basadas en normas y referidas a criterio.

El capítulo 6 se refiere al tema de los cuestionarios y se tratan los aspectos siguientes: definición, propósito y clasificación de los cuestionarios y pasos para el diseño de un buen cuestionario. Finalmente, el capítulo 7, está dedicado al tema del Análisis de Factores (comúnmente identificado en el idioma español como análisis factorial), en el cual se plantea un resumen comprensivo de esta técnica analítica multivariable que incluye desde su historia hasta los aspectos prácticos de su utilización y aplicación en el proceso del desarrollo de los instrumentos de medición en ciencias sociales.

La segunda parte de la obra está dedicada a facilitar información a los usuarios acerca del uso de algunas técnicas de recolección y análisis de información y datos cualitativos. El capítulo 8 se refiere a la observación participante y a la entrevista en profundidad; comprende los aspectos siguientes: definición de observación, tipos de observación, la observación participante: preparación, entrada y salida del escenario, observación no-participante; aplicaciones a la investigación educativa. Entrevista en profundidad: definición, preparación, desarrollo y cierre; aplicaciones en el contexto educativo.

En el capítulo 9 se abordan las técnicas del grupo de enfoque y la observación flotante; se desarrollan los aspectos siguientes: definición de grupo focal, antecedentes, aplicaciones, procedimientos: planificación, ejecución y cierre. La observación flotante: definición, contextualización del escenario, enfoques en el análisis de la información: antropológico, lingüístico y psicopedagógico.

Finalmente, en el capítulo 10, se describen los procedimientos para el análisis de datos cualitativos; incluye los aspectos siguientes: definición, fases de análisis: descripción y teorización; y el uso de programas electrónicos para el análisis cualitativo.

Esperamos que esta tercera edición de la obra, al igual que las dos anteriores, satisfaga las expectativas de los usuarios, particularmente de los estudiantes de pregrado y postgrado en proceso de elaborar su trabajo o tesis de grado y de los profesores que trabajan en las cátedras de investigación en educación y/o psicología o que se desempeñan como asesores, tutores, directores o supervisores de investigación académica.

# Introducción

La investigación científica es un proceso sistemático de aproximación metódica a la realidad con el propósito de conocerla, comprenderla, describirla, explicarla o transformarla. No es sinónimo de la acción simple de recoger datos y/o información, sin planificación, acerca de un hecho o fenómeno determinado. Supone una concepción de la forma y naturaleza de dicha realidad u objeto de estudio (ontología); una visión de cómo se produce el conocimiento, en términos de la relación asumida entre el sujeto cognoscente y el objeto cognoscible (epistemología) y una manera operacional de indagar sobre el objeto de estudio (método). Es decir, investigar implica abordar la realidad desde una perspectiva onto-epistémica y metodológica para la producción de conocimiento. En este contexto, surge la importancia de los instrumentos y técnicas de investigación -tema central de esta obra- como los medios que hacen posible tener un contacto lo más cercano posible con la realidad objeto de estudio. En consecuencia, con esta introducción se establecen las bases teóricas necesarias que fundamentan y justifican el uso de tales procedimientos de investigación, a partir del paradigma epistémico adoptado.

Etimológicamente, el vocablo paradigma (Del latín *paradigma*, y éste del griego παράδειγμα), significa modelo, patrón o ejemplo (RAE, 2001). Así, en sentido general, un paradigma puede ser definido como “un conjunto de reglas y disposiciones (escritas o no) que hace dos cosas: (a) establece o define límites; y (b) indica como comportarse dentro de los límites para tener

éxito” (Barker, 1995, p. 35). Esta noción genérica de lo que es un paradigma esta implícita también, de alguna manera, en la concepción de los paradigmas de la ciencia. En este sentido, Kuhn (1978) ha señalado que “los paradigmas científicos son ejemplos aceptados de la práctica científica actual, ejemplos que combinan ley, teoría, aplicación e instrumentación y proporcionan modelos a partir de los cuales se manifiestan las tradiciones coherentes particulares de la investigación científica [...] Los hombres cuya investigación se basa en paradigmas compartidos están sujetos a las mismas reglas y patrones en la práctica científica” (p. 10).

Históricamente han existido diferentes maneras de abordar el estudio de la realidad social, tanto desde el punto de vista teórico-filosófico como metodológico. En este sentido, coexisten contemporáneamente tres paradigmas de investigación en las ciencias sociales como son: el positivista, el interpretativo y el crítico, los cuales se describen brevemente a continuación.

## Paradigma Positivista

El paradigma positivista se fundamenta en las ideas filosóficas de Augusto Comte, quien enfatizó la observación y la razón como medio de comprender la conducta humana. Consideraba que el conocimiento era el resultado de la experiencia y podía ser obtenido mediante la observación y el experimento. Adoptó el método científico como medio de generación de conocimientos. De allí que este paradigma tenga que ser entendido en el marco de los principios y supuestos de la ciencia, como son: el determinismo, el empirismo, la parsimonia y la generalización.

El determinismo supone que los eventos o fenómenos tienen ciertas causas y el estudio de la relación causa-efecto es necesario para su predicción y control. El empirismo implica la recolección de evidencias verificables de la realidad objetiva, como soporte a las hipótesis y teorías. La parsimonia se refiere a la explicación de un fenómeno en la manera más sencilla y económica posible; mientras que la generalización es el proceso de proyectar el resultado de las observaciones particulares de un fenómeno al universo como un todo. De acuerdo con estos supuestos, la ciencia busca integrar y sistematizar resultados dentro de patrones conceptuales o teorías explicativas que pueden ser revocables.

Algunas de las características del paradigma positivista son las siguientes: (a) asume la existencia de un mundo material externo que puede ser descrito objetivamente, sin referencia al observador; (b) esta realidad comprende todo lo que existe dentro de las coordenadas: tiempo, espacio y masa; (c) tiene como objeto de estudio los hechos y fenómenos que pueden ser observables y verificables; (d) enfatiza la observación experimental, la

medición objetiva y la generalización de los resultados; (e) la producción del conocimiento es el resultado de la interacción sujeto-objeto mediada por el método; (f) utiliza instrumentos y procedimientos estandarizados; (g) enfatiza la objetividad, replicabilidad y generalización del conocimiento; (h) estudia las causas que explican los fenómenos y confronta la teoría con la praxis; (i) establece relaciones generalizables entre variables; asume una neutralidad valorativa; y (j) utiliza el método hipotético-deductivo.

La crisis del positivismo como paradigma epistémico dominante durante varios siglos fue la consecuencia de los cambios revolucionarios ocurridos en algunos conceptos básicos de la física durante el primer tercio del siglo pasado. Martínez (2004) señala los principales autores y sus contribuciones científicas que hicieron posible tal revolución, ellos fueron: Einstein (relativiza los conceptos de espacio y tiempo); Heisenberg (introduce el principio de incertidumbre); Pauli (formula el principio de exclusión) Niels Bohr (establece el principio de complementariedad); Planck, Schrodinger y otros (descubren con la mecánica cuántica un conjunto de relaciones que gobiernan el mundo subatómico). El mismo Martínez (2004) considera que estos hechos son interpretados por muchos autores como el fin del realismo ingenuo, propio de la epistemología positivista, y como la bancarrota del determinismo físico. Los principios se aplican a partículas y acontecimientos microscópicos; pero estos acontecimientos tan pequeños no son, en modo alguno, insignificantes. Son precisamente el tipo de acontecimientos que se producen en los nervios y en el cerebro, como también en los genes y, en general, son la base que constituye toda materia del cosmos y todo tipo de movimiento y forma de energía.

Estos nuevos conocimientos en la microfísica cambiaron el modo de pensar acerca de la ciencia, lo que permitió reconocer que los fenómenos de

las ciencias naturales y las ciencias sociales son de naturaleza diferente y, por tanto, deben ser estudiados desde distintas perspectivas. Emerge así y se consolida el paradigma naturalista-interpretativo como una expresión del modo de conocer postpositivista.

## Paradigma Interpretativo

El paradigma naturalista-interpretativo se fundamenta principalmente en el pensamiento fenomenológico (Husserl, 1962); la hermenéutica (Dilthey, 1972); y en el interaccionismo simbólico (Blumer, 1982). Este nuevo paradigma asume, desde un punto de vista ontológico, que la realidad objetiva, como tal, no existe (a diferencia de la concepción positivista) sino que es una construcción social, resultado de la interacción humana. En consecuencia, se trata de una realidad dinámica y singular, determinada histórica y socioculturalmente. En el plano epistemológico, se asume la producción de conocimiento a partir del mundo interior de los actores sociales, donde el sujeto y el objeto no son separables puesto que forman parte del mismo fenómeno bajo estudio, por lo que se influyen mutuamente; se reivindica la subjetividad como fuente de conocimiento; la investigación enfatiza la comprensión e interpretación de la realidad y se aboca al estudio de casos.

Asimismo, se niega la posibilidad de una relación de causa-efecto entre los fenómenos sociales ya que estos se encuentran en una situación de influencia mutua, por lo que no resulta factible distinguir entre causas y efectos. Igualmente, frente a la suposición positivista del desarrollo de un cuerpo nomotético de conocimiento bajo la forma de generalizaciones universales, plantea la construcción de cuerpo ideográfico de conocimientos capaz de describir el objeto de investigación.

Desde el punto de vista metodológico, la investigación cualitativa enfatiza el estudio de los problemas de la vida cotidiana y se fundamenta en los procesos de pensamientos de los actores acerca de las acciones,

interacciones y transacciones en las que se involucran en distintos contextos socioculturales. Destaca el papel de los valores en la investigación social. En este sentido, considera que cualquier tipo de actividad investigadora está influida por: (a) la intención del investigador; (b) la elección del paradigma asumido; (c) la orientación de la teoría que fundamenta la investigación, guía la recolección y análisis de la información, y la interpretación de los hallazgos; y (d) los valores que forman parte del contexto en el que se desarrolla el trabajo.

## Paradigma Sociocrítico Crítico

Este modelo de investigación social surge como una reacción tanto al paradigma interpretativo como al positivismo. En el primer caso, se cuestiona el alcance pasivo de la investigación en el sentido que se agota en la comprensión, descripción e interpretación, siendo que la investigación social debe ser transformadora. Es decir, se investiga para cambiar la estructura de la realidad. En cuanto al positivismo, se critica su racionalidad instrumental y técnica exigiéndose la necesidad de una racionalidad substantiva que incluya los juicios, los valores y los intereses de la humanidad. En otras palabras, esta perspectiva pretende superar el reduccionismo positivista y el conservadurismo interpretativo, admitiendo la posibilidad de una ciencia social que no sea ni puramente empírica ni solo interpretativa.

La teoría crítica se origina en la llamada escuela de Frankfurt, integrada por un grupo de filósofos neomarxistas ([Adorno](#), [Marcuse](#), [Habermas](#), [Schweppenhäuser](#), [Fromm](#), [Wellmer](#) y [Honneth](#), entre otros) unidos en torno a la necesidad de superar el planteamiento de la razón instrumental tradicional por una razón subjetiva, fundamentada en los sentimientos humanos. Esta teoría asume que “el conocimiento no es una simple reproducción conceptual de los datos objetivos de la realidad, sino una auténtica formación y constitución de la misma. Se opone radicalmente a la idea de teoría pura que supone una separación entre el sujeto que contempla y la verdad contemplada, e insiste en un conocimiento que está mediado por la experiencia, por las praxis concretas de una época, como por los intereses teóricos y extra-teóricos que se mueven al interior de las mismas”.

Entre sus principales características están: (a) asume una ontología

realista crítica, fundamentada en la ideología, y una epistemología subjetivista, en el sentido de que los valores son fundamentales para la investigación; (b) propone una metodología dialógica, participativa y transformativa, orientada hacia la emancipación del hombre; dicha metodología asume dos modalidades, a saber: la investigación – acción y la crítica ideológica; y (c) considera fundamental la relación entre teoría y praxis, la cual constituye el criterio básico que le permite diferenciarse de los distintos paradigmas o tradiciones de la investigación.

El paradigma crítico introduce la ideología de forma explícita de la autorreflexión crítica e los procesos del conocimiento. Tiene como finalidad la transformación de la estructura de las relaciones sociales y dar respuesta a determinados problemas generados por éstas. Sus principios son: (a) conocer y comprender la realidad como praxis; (b) unir teoría y práctica (conocimiento, acción y valores); (c) orientar el conocimiento a emancipar y liberar al hombre; y (d) implicar al docente a partir de la autorreflexión.

## **Enfoques y Métodos de Investigación**

Hasta aquí nos hemos referido a los tres principales paradigmas de investigación que coexisten en las ciencias sociales, como son el positivista, el interpretativo y el sociocrítico. No obstante, por razones de alcance y propósitos de esta obra nos referiremos, en adelante, solamente a los dos primeros, en términos de sus enfoques y métodos asociados.

## Enfoque Cuantitativo

El enfoque cuantitativo representa el conjunto de opciones metodológicas a través de las cuales se expresa el paradigma positivista en la investigación de los hechos sociales. Algunas de sus características son: (a) los hechos son estudiados en términos de variables; (b) las variables son medidas mediante el empleo de instrumentos válidos y confiables; (c) los datos son analizados mediante técnicas estadísticas; (d) los resultados son organizados en cuadros y gráficos; (e) es secuencial y probatorio; (f) los procesos siguen un orden riguroso; (g) las decisiones críticas se toman antes de recolectar los datos; y (h) busca describir las situaciones estudiadas, explicar los hechos en términos de relación de causa-efecto y predecir situaciones futuras a partir de los datos obtenidos. Los principales métodos que utiliza la investigación cuantitativa son: experimental, descriptivo y correlacional.

### **El Método Experimental.**

La investigación experimental consiste en someter a verificación una situación novedosa (variable independiente) para estudiar el efecto que produce sobre un hecho hipotéticamente relacionado (variable dependiente), bajo condiciones de control (variables intervinientes). Sus principales características son las siguientes: (a) la equivalencia estadística de sujetos en diversos grupos formados aleatoriamente; (b) comparación de dos o más grupos o conjunto de condiciones; (c) manipulación directa de la variable independiente; (d) medición de cada variable dependiente; (e) uso de la estadística inferencial; y (f) utilización de un diseño que permita la maximizar la varianza experimental, minimizar la varianza de error y controlar la varianza extraña.

Una investigación experimental se desarrolla mediante la puesta en práctica de las siguientes fases: (a) identificar un problema de investigación; (b) formular las hipótesis correspondientes; (c) seleccionar e instrumentar un diseño experimental apropiado; (d) recolectar la información y analizar los datos; y (e) obtener las conclusiones. Dentro de esta categoría de investigación existen varias opciones o modalidades, como son: (a) la investigación experimental propiamente dicha, en la que se satisfacen todas las exigencias formales de este método (por ejemplo, el diseño de grupo control con preprueba y posprueba); (b) la investigación cuasi-experimental, en la que solo se cubren parcialmente las exigencias formales del método experimental (por ejemplo, el diseño de grupo control no equivalente); y (c) la investigación pre-experimental, la cual tiene limitadas condiciones de validez, ya que no satisface las exigencias mínimas del método de investigación experimental; se incluye dentro de las opciones como una indicación del tipo de modalidad de investigación experimental que se debe evitar (por ejemplo, el diseño de estudio de caso con una sola medición).

### **El Método Descriptivo.**

Desde el punto de vista científico, describir implica decir algo acerca de un hecho a partir de la medición de sus variables más relevantes. Por lo tanto, la investigación descriptiva tiene como propósito determinar las características de un hecho o fenómeno. Es decir, nos indica cómo es y cómo se manifiesta una determinada situación bajo estudio. Para Danhke (1989) “los estudios descriptivos buscan especificar las propiedades importantes de personas, grupos, comunidades o cualquier otro fenómeno que sea sometido a análisis”. Existen diferentes modalidades de estudios descriptivos, como son: la investigación por encuesta, el estudio de casos, el estudio de tendencia, la investigación transversal y la investigación longitudinal, entre otras.

Para realizar un estudio descriptivo se requiere seguir los pasos siguientes: (a) definir el problema o la situación objeto de estudio; (b) seleccionar una muestra aleatoria de sujetos de la población involucrada; (c) aplicar instrumentos de medición (por ejemplo cuestionarios, escalas de estimación, pruebas); (d) transcribir y analizar los datos; (e) interpretar los resultados y obtener conclusiones. Puesto que los estudios descriptivos se realizan con base en muestras estadísticas, lo más importante de este tipo de investigaciones esta en la posibilidad de generalización que se puede hacer a la población estudiada, a partir de los resultados obtenidos en la muestra. Ejemplos de estudios descriptivos son los censos de población que se realizan en los países cada cierto tiempo para conocer las características de la población en cuanto a variables tales como: edad, sexo, estado civil, educación, vivienda, empleo y nivel de ingreso entre otras. Tambien se incluyen en este grupo, los estudios de tendencias en la intención del voto en los procesos electorales.

### **El Método Correlacional.**

Consiste en seleccionar un conjunto de variables sobre un problema de interés, medirlas y determinar el grado de relación que existe entre ellas. Esta relación podría asumir varias formas, una de ellas es el tipo de correlación entre dos variables. Por ejemplo, un investigador podría estar interesado en responder la pregunta siguiente ¿Qué relación existe entre la capacidad intelectual y el rendimiento académico de los estudiantes? Para ello, mide las variables con sendas pruebas de inteligencia y desempeño académico respectivamente, utiliza una técnica estadística apropiada (por ejemplo, la correlacion producto momento de Pearson) y obtiene un índice de correlación que al ser interpretado nos indicará si el grado de relación entre las dos variables existe o no; y en caso de ser cierta, nos dirá si la relación es positiva

o negativa, si su magnitud es baja, moderada o alta y si el coeficiente obtenido es estadísticamente significativo; es decir, si valor en el contexto poblacional es diferente a cero (0).

La relación se podría plantear también entre tres o más variables; por ejemplo, si incluyéramos la variable motivación en la pregunta anterior, nos plantearíamos ¿Cuál es la relación entre la motivación, la capacidad intelectual y el rendimiento académico de los estudiantes? Por otra parte, los estudios correlacionales se podrían realizar con base en modelos estadísticos *univariados*, cuando la relación se establece entre una variable independiente y otra dependiente, o *multivariados*, cuando la relación se establece entre varias variables independientes y varias dependientes (por ejemplo, la correlación Canónica).

Los estudios correlacionales, además de permitirnos conocer las relaciones entre las variables, hacen posible hacer predicciones sobre el comportamiento de un grupo de sujetos en una determinada variable a partir de las puntuaciones obtenidas por los sujetos en otra variable o variables relacionadas. Por ejemplo, si sabemos que la capacidad intelectual correlaciona alto con el desempeño académico, podríamos utilizar esa información para pronosticar el rendimiento futuro de un estudiante a partir de la puntuación que obtuvo en un test de capacidad intelectual.

## Enfoque Cualitativo

El enfoque cualitativo representa el conjunto de opciones metodológicas a través de las cuales se expresa el paradigma interpretativo en la investigación de los fenómenos sociales. Provee información acerca del lado humano de los hechos sociales. En este sentido se enfoca, desde una perspectiva subjetivista, al estudio de comportamientos, creencias, opiniones, emociones, actitud, valores, sentimientos y la relación entre las personas. Este enfoque es efectivo también en la identificación de factores intangibles tales como: normas sociales, estatus socioeconómico, género, etnicidad y religión, cuyo rol en el proceso de investigación pudiera no ser evidente.

Este enfoque utiliza diferentes métodos de indagación, entre los que se encuentran: el etnográfico, fenomenológico, hermenéutico, interaccionismo simbólico, historia de vida y el método comparativo constante, entre otros. Asimismo, emplea diferentes técnicas para la obtención de la información acerca de los fenómenos estudiados, entre ellas: la observación participante, entrevista en profundidad y el grupo de enfoque, entre otras. En definitiva se trata de obtener información significativa, desde la perspectiva de los actores sociales, mediante diferentes técnicas con el propósito de analizarla, codificarla, categorizarla y relacionarla a fin de comprenderla, describirla e interpretarla. A continuación se describen algunos de los métodos de indagación que utiliza el enfoque cualitativo.

### **El método etnográfico.**

Etimológicamente, el vocablo etnografía proviene del griego, *ethnos*, εθνος, que significa “tribu, pueblo y *grapho*, γραφω, que quiere decir *descripción*; literalmente “descripción de los pueblos”. A través de este método se

pretende revelar los significados que sustentan las acciones, interacciones y transacciones que constituyen la realidad social del grupo estudiado. Es probablemente el método más antiguo y popular utilizado originalmente por los antropólogos culturales para estudiar el modo de vida de una unidad social determinada, pudiendo ser ésta una tribu indígena, un barrio o una comunidad urbana, una familia o una empresa.

Más contemporáneamente, el método etnográfico ha sido utilizado por otros investigadores como sociólogos, psicólogos sociales y educadores para estudiar los procesos humanos que ocurren en grupos concretos tales como los actores de un hospital, una cárcel, una escuela o un aula de clase. Recientemente, con el auge de las tecnologías de la información y la comunicación y sus aplicaciones al contexto educativo se han realizado estudios de las comunidades de aprendizaje que se producen en los entornos de educativos virtuales mediante el método de la etnografía virtual o ciber-etnografía (**Escobar**, 1994).

De acuerdo con Martínez (1989), este método se apoya en la convicción de que las tradiciones, roles, valores y normas del ambiente en que se vive se van internalizando poco a poco y generan regularidades que pueden explicar la conducta individual y grupal en forma adecuada. Su puesta en práctica supone llevar a cabo, durante largos períodos de tiempo, una observación directa de la vida diaria de la unidad social estudiada y la realización de entrevistas en profundidad a sus miembros que permitan la recogida de información detallada sobre el comportamiento social, registrando una imagen realista y fiel del grupo estudiado. Esta información, conjuntamente con determinados referentes teóricos ayudara a comprender e interpretar los procesos humanos que se dan en dicho grupo.

### **El método fenomenológico.**

La fenomenología es una corriente filosófica creada por Husserl (1859-1938) en la primera mitad del siglo pasado. Consiste en el examen sistemático de la subjetividad; su meta es el estudio del mundo tal como se nos presenta en y a través de la conciencia; es decir, es la investigación acerca de los fenómenos tal y como son experimentada, vivida y percibida por el individuo. Su propósito es indagar sobre cómo las personas comprenden los significados de los eventos (Patton, 1990). Sus principales características son: (a) la primacía que otorga a la experiencia subjetiva inmediata como base del conocimiento; (b) el estudio de los fenómenos desde la perspectiva de los sujetos, teniendo en cuenta su marco referencial; y (c) el interés por conocer cómo las personas experimentan e interpretan el mundo social que construyen en interacción con otros.

Martínez (1989) considera que el método fenomenológico se centra en esas realidades vivenciales que son poco comunicables, pero que son determinantes para la comprensión de la vida psíquica de cada persona. Asimismo, plantea que la aplicación de este método requiere que el investigador inicie el estudio sin hipótesis previa, tratando de reducir al mínimo la influencia de las propias teorías, ideas e intereses y hacer un gran esfuerzo por captar toda la realidad que se presenta de manera vivencial a nuestra conciencia. En este sentido, es importante que el investigador oiga desprejuiciada y detalladamente muchos casos similares, describir minuciosamente cada uno de ellos y elaborar una estructura común representativa de esas experiencias vivenciales.

### **El método hermenéutico.**

El término hermenéutico proviene del verbo griego *ῥηνεύειν* (*jermeneueien*) que significa *interpretar, declarar, anunciar, esclarecer* y, por último, *traducir*. Dilthey (citado por Martínez, 1989) define la hermenéutica como

“el proceso por medio del cual conocemos la vida psíquica con la ayuda de signos sensibles que son su manifestación”. Es decir, tiene como misión descubrir los significados de las cosas, interpretar lo mejor posible las palabras, los escritos, los textos, los gestos y, en general, el comportamiento humano, así como cualquier acto u obra suya, pero conservando su singularidad en el contexto del que forma parte. En este sentido, la hermenéutica se convierte en un método de sistematización de procedimientos formales que permiten la correcta interpretación y comprensión del significado de las acciones humanas, las cuales no siempre resultan evidentes.

Dilthey propuso la técnica del círculo hermenéutico como una herramienta heurística

que permite la interpretación de la información a partir de un análisis de la relación dialéctica que se establece entre el todo (visto como estructura) y las partes que lo integran. Particularmente relevante para dicho autor es la noción de estructura por cuanto permite captar en una totalidad la coherencia de los diversos elementos, en función esencialmente de su finalidad consciente e inconsciente; y busca una ciencia de las realidades humanas que produzca un conocimiento cierto y objetivo, es decir, verificable de

manera intersubjetiva, consciente de que hay grados de verdad y que a ella sólo se llega por aproximación (Martínez, 1989).

La técnica del círculo hermenéutico se fundamenta en los principios siguientes: (a) el todo explica la parte y la parte explica el todo; (b) es preciso comprender los prejuicios que tenía el autor del texto, lo que parecía obvio en la mentalidad de su tiempo (horizonte del autor); (c) siempre interpretamos un texto desde nuestros propios prejuicios (horizonte del lector); (d) para interpretar se debe poner en diálogo los prejuicios del autor y los del lector

(fusión de horizontes); (e) el texto “forma” al lector (*Wirkungsgeschichte*) y el lector “reforma” el texto; (f) comprender un texto es también comprenderse a sí mismo; y (g) la comprensión de un texto nunca se termina (Wikispaces, 2013). El proceso de investigación hermenéutica contempla las fases siguientes:

1. *Determinar el objetivo que se persigue.* Es necesario precisar que se va a estudiar y por qué. En lo posible la intención de la investigación debería expresarse en forma de una pregunta.
2. *Elección de los informantes a estudiar.* Aquí es importante resolver el problema de cuántos y quiénes son los informantes (características personales y socioculturales), lo cual siempre va a depender del propósito de la investigación y de la intencionalidad y conocimiento del investigador.
3. *Selección de los instrumentos y procedimientos de aplicación.* En este aspecto, Martínez (1989) recomienda tomar en cuenta dos recomendaciones: (a) los instrumentos y técnicas deben ser no intrusivos en el sentido de no alterar la naturaleza del objeto de estudio; debe darse un proceso ecológico de indagación; y (b) los procedimientos utilizados deben permitir realizar las observaciones repetidas veces. Una de las técnicas que se recomienda en este tipo de investigación es la entrevista semiestructurada o dialogo coloquial.
4. *Análisis e Interpretación.* La estrategia incluye las fases siguientes: (a) establecer las unidades de análisis las cuales pueden estar compuestas por palabras, frases, o párrafos, que tienen o exhiben una *idea central unitaria*, y pueden estar subsumidas en otras más amplias; y (b) la *categorización* de cada unidad de análisis, la cual consiste en ponerle un

nombre breve o etiqueta que sintetice el significado de la unidad. Varias categorías pueden relacionarse entre sí de diversas formas, constituyendo una estructura, lo cual permitirá al investigador identificar la realidad subyacente en los “datos” que se están analizando.

### **El interaccionismo simbólico.**

Este método fue propuesto por Blumer (1982). Se fundamenta en la filosofía del pragmatismo y propone como objeto de estudio la interacción humana desde una perspectiva cualitativa. En el mismo se privilegia la acción como interacción comunicativa, como proceso interpersonal y al mismo tiempo auto-reflexivo. Tiene como propósito explicar la interacción en el individuo y en los grupos, y esta explicación se construye desde una perspectiva evolutiva, es decir, histórica.

Los postulados básicos del interaccionismo simbólico, según Blumer (op. cit), son los siguientes:

1. Las personas actúan sobre los objetos de su mundo e interactúan con otras personas a partir de los significados que los objetos y las personas tienen para ellas. Es decir, a partir de los símbolos. El símbolo permite, además, trascender el ámbito del estímulo sensorial y de lo inmediato, ampliar la percepción del entorno, incrementar la capacidad de resolución de problemas y facilitar la imaginación y la fantasía.
2. Los significados son producto de la interacción social, principalmente la comunicación, que se convierte en esencial, tanto en la constitución del individuo como en (y debido a) la producción social de sentido. El signo

es el objeto material que desencadena el significado, y el significado, el indicador social que interviene en la construcción de la conducta.

3. Las personas seleccionan, organizan, reproducen y transforman los significados en los procesos interpretativos en función de sus expectativas y propósitos.
4. El individuo se constituye en la interacción social (formación del yo social autoconsciente); considera que no es posible entender el yo sin el otro ni a la inversa, y que los grupos y la sociedad se constituyen sobre la base de las interacciones simbólicas de los individuos al tiempo que las hacen posibles. Desde el punto de vista metodológico, el interaccionismo simbólico utiliza preferentemente el diseño de investigación longitudinal, estudio de casos comparativos, utilizando diferentes fuentes de información, y la técnica de la triangulación de fuentes, utiliza la observación participante y la información documental, combinando el procedimiento empírico/inductivo con entrevistas entre semi-estructuradas y mixtas.

### **El método biográfico.**

Este método pertenece a una categoría de indagación cualitativa más amplia como lo es la investigación narrativa entendida como una forma particular de reconstrucción de la experiencia por la que, mediante de un proceso reflexivo, se da significado a lo sucedido o vivido (Ricoeur, 1995). El método biográfico se conoce en la literatura también como historia de vida y consiste en un relato autobiográfico obtenido por el investigador mediante entrevistas sucesivas en las que el objetivo es mostrar el testimonio subjetivo de una persona en la que se recojan tanto los acontecimientos como las valoraciones que dicha persona hace de su propia existencia. El investigador es sólo el

inductor de la narración, su transcriptor y, también, el encargado de retocar el texto (Pujadas, 1992).

En consecuencia, el proceso de obtener los relatos del informante no es sinónimo de recolección de datos en el sentido tradicional de la investigación, sino que se trata más bien de asistir y participar en la elaboración de una memoria que quiere transmitirse a partir de la demanda del investigador. De allí que Santamarina y Marinas (1995) consideren que la historia de vida no es sólo transmisión sino una construcción en la que participa el propio investigador. A través de la historia de vida el investigador intenta trascender el alcance individual particular de la información biográfica obtenida para buscar comprender el medio social, los procesos sociales a partir de las experiencias de una persona, de un grupo o una organización.

El procedimiento seguido en la historia de vida comprende cuatro fases a saber: la problemática, la recolección de información, la transcripción y el análisis e interpretación. En relación con la primera fase, los especialistas recomiendan la elaboración de un problema a investigar, la formulación de hipótesis y la puesta en marcha de proposiciones de interpretación. En cuanto a la segunda, es muy importante el desarrollo previo de un *raportt* apropiado que genere confianza en el entrevistado a objeto de que pueda abrirse a la conversación. Asimismo, es necesario establecer toda una logística para llevar a cabo el proceso de obtención de la información; Durante esta fase es importante tomar decisiones sobre preguntas tales como: donde entrevistar, cuantas entrevista hacer y sobre cuál(es) tema(s), entre otras. Algunos investigadores consideran que se debe tener una programación flexible donde las entrevistas se realicen a ritmos específicos sin mucha regularidad, atendiendo más bien a las posibilidades del entrevistado.

En cuanto a la tercera fase (transcripción) es el momento en que el

investigador reproduce la entrevista realizada ya sea de memoria, escrita o en audiograbadora e inicia el proceso de dar forma al discurso del informante. Finalmente, en la fase de análisis e interpretación, aun cuando la misma está presente a través de todo el proceso de investigación, en lo específico se trata de identificar los aspectos relevantes de la información recopilada, precisar sus relaciones y buscar sentido a la estructura emergente; para ello se pueden utilizar las técnicas del análisis de contenido o análisis del discurso u otra que se considere apropiada. En cuanto a las aplicaciones de la historia de vida al contexto educativo, Colas (1997) considera que a través de ella:

Se puede llegar a averiguar qué conocen los profesores de la enseñanza, cómo está organizado el conocimiento y cómo cambia su conocimiento a través de la experiencia. La historia puede reflejar el discurso dominante de los profesores. (...) Las Historias de los profesores constituyen una importante línea de investigación sobre la formación del profesorado. Su aplicación resulta también oportuna para el estudio de “las voces” en el aula (p. 282).

### **El método comparativo constante.**

Este método (MCC) está basado en una propuesta formulada por Glaser y Strauss (1967), ampliada posteriormente por Strauss y Corbin (1998), como una alternativa para el estudio de la realidad social, enfocándose hacia la elaboración teórica a partir de la reflexión crítica sobre los datos e información obtenidos en el proceso de investigación. Es decir, el propósito de la investigación que proponen estos autores es el de crear una teoría fundamentada en los datos, que tuviese un doble sentido: comprender y explicar la realidad, al mismo tiempo que ser una guía para la acción (aplicación); es decir, que tuviese sentido tanto para el científico social como

para la gente común. De allí que con el tiempo el método ha ido adoptando el nombre de teoría fundamentada.

El procedimiento del MCC es como sigue: (a) se selecciona un problema y se formula la pregunta de investigación; (b) se inicia la recolección de datos e información; (c) se procede al micro análisis de material obtenido; (d) se comienza el proceso de codificación (abierta, axial, selectiva) de los datos; (e) se formulan categorías relevantes; (f) se establecen relaciones entre las categorías encontradas, identifican conceptos nuevos y se inicia el proceso de teorización. Strauss y Corbin (2002) señalan que “A medida que los conceptos y las relaciones emergen de los datos por medio del análisis cualitativo, el investigador puede utilizar esa información para decidir dónde y cómo ir consiguiendo información adicional para el avance posterior de la teoría” (p. 37).

El MMC utiliza el procedimiento del muestreo teórico como estrategia para la obtención sistemática de datos e información en el proceso de la construcción de la teoría fundamentada. Strauss y Corbin (2002) lo definen como “recolección de datos guiada por los conceptos derivados de la teoría que se esta construyendo y basada en el concepto de hacer comparaciones, cuyo propósito es acudir a lugares, personas o acontecimientos que maximicen las posibilidades de descubrir variaciones entre los conceptos y que hagan más densas las categorías en términos de sus propiedades y dimensiones” (p. 219).

## Enfoque Multimétodo

El enfoque multimétodo (EMM) puede ser entendido como una estrategia de investigación en la que se utilizan dos o más procedimientos para la indagación sobre un mismo fenómeno u objeto de estudio a través de los diferentes momentos del proceso de investigación, como son: (a) la concepción de la investigación y preguntas asociadas; (b) la selección del enfoque de investigación; (c) el levantamiento y análisis de la información; y (d) la interpretación de los resultados e inferencia.

Quienes optan por utilizar el EMM en la investigación social esgrimen diferentes argumentos. Personalmente, considero que una de las razones para utilizar esta opción de investigación tiene que ver con su mayor flexibilidad para adaptarse a las demandas de comprensión y explicación de una realidad, como la actual, caracterizada por su multidimensionalidad y complejidad, todo lo cual permite ir más allá en el conocimiento del objeto de lo que podrían aportar, en forma independiente, cada uno los componentes individuales de dicho enfoque.

Otros autores hacen consideraciones como las siguientes: esta modalidad de indagación contribuye a mejorar los procesos y productos de investigación (Tashakkori y Teddlie, 2003); su combinación produce información cuantificable y contextual (Kaplan y Duchon, 1988); recolectar diferentes tipos de datos, por diferentes métodos y fuentes provee de una información de un mayor alcance, lo que podría resultar en un cuadro más completo del objeto de estudio en comparación con lo que se podría lograr utilizando separadamente cada método en particular (Bonoma, 1985); cada uno de los métodos tiene fortalezas y debilidades y el EMM, conscientemente

utilizado permite potenciar los aspectos favorable de cada uno de ellos, al mismo tiempo que compensa sus debilidades. Es decir, ninguno de los métodos está libre de prejuicios, sólo puede llegar a la verdad subyacente mediante el empleo de múltiples técnicas con las que el investigador efectuará las correspondientes triangulaciones (Cook y Reichardt, 1986).

La legitimidad científica de la integración metodológica para el abordaje de lo social es probablemente uno de los aspectos más polémicos, especialmente cuando se la mira desde las perspectivas onto-epistemológicas de los enfoques cuantitativo y cualitativo. Particularmente, considero que tal enfoque de investigación es legítimo desde el punto de vista de asumir una realidad como multidimensional y compleja, todo lo cual supone un abordaje metodológico que trascienda las visiones parciales de los dos enfoques. Sin embargo, en la práctica, la validez científica de los diferentes diseños multimétodo depende del conocimiento que el investigador tenga de los supuestos metateóricos que asumen tanto el paradigma positivista como el naturalista-interpretativo, a fin evitar contradicciones a este nivel y en aras de lograr una integración metodológica armónica que permita tanto la comprensión como la explicación de los fenómenos estudiados.

Para concluir, al cierre de esta Introducción, podríamos señalar que en toda investigación social esta implícita la noción de paradigma, como metateoría que orienta el quehacer investigativo, al cual se subordinan los medios –instrumentos y técnicas- que permiten el abordaje de la realidad objeto de estudio; es decir, los instrumentos y técnicas como procedimientos operacionales del proceso investigativo deben ser seleccionados o diseñados en correspondencia con el paradigma epistémico que le sirve de referencia. De allí que el investigador deba estar consciente de esta relación necesaria a objeto de evitar posibles contradicciones que dificulten la interpretación de

los resultados obtenidos en el proceso de indagación. En caso del EMM debe quedar claro que aun cuando no existe ninguna posibilidad de integrar los enfoques cuantitativo y cualitativo a nivel metateorico ya que parten de fundamentaciones onto-epistémicas irreconciliables, si es posible tal integracion en nivel empírico donde el objeto de estudio puede ser abordado complementariamente con diferentes instrumentos y técnicas de obtención y análisis de datos.

PART 1

PARTE I INSTRUMENTOS DE  
INVESTIGACION CUANTITATIVA

# Instrumentos de Medición

En el contexto de las ciencias de la educación, los instrumentos de medición son procedimientos sistemáticos y estandarizados que permiten observar la conducta humana, a fin de hacer inferencias sobre determinados constructos, rasgos, dimensiones o atributos. En el caso de la investigación educativa, Best (1973) ha definido los instrumentos como “aquellos objetos materiales que nos permiten adquirir y analizar datos mediante los cuales pueden ser comprobadas las hipótesis de la investigación” (p. 133).

De acuerdo con la formulación clásica de Stevens (1951), la medición se entiende como el proceso de asignar numerales a los objetos o eventos con base en ciertas reglas. Clarifiquemos esta definición con un ejemplo. Supongamos que queremos medir hasta qué punto un grupo de estudiantes, de noveno grado de educación básica, tiene un dominio alto, promedio o bajo de los objetivos de matemática. Asuma, además, que el grupo fue sometido a una prueba de rendimiento académico en dicha asignatura, la cual fue calificada con base en una escala del 1 al 10, tomando en cuenta los siguientes criterios: dominio alto, puntajes entre 8 y 10; dominio promedio, puntajes entre 5 y 7; y dominio bajo, puntajes entre 1 y 4.

En el caso anterior, el objeto de medición fue el dominio de la matemática, por parte de un grupo de alumnos de noveno grado de educación básica. Los numerales, se referían a los valores (1 a 10) de la escala utilizada; mientras que las reglas empleadas aludían a los criterios de clasificación utilizados para ubicar a cada uno de los miembros del grupo en la categoría

correspondiente (alto, promedio, bajo), de acuerdo con el puntaje obtenido en la prueba de rendimiento.

En esta definición de medición, es importante aclarar; primero, que la misma se refiere tanto a objetos físicos (longitud, peso, superficie o capacidad de un objeto) como psicológicos (constructos o variables orgánicas, de observación indirecta); segundo, que el concepto de numeral no implica necesariamente cantidad. En el contexto de la medición, un número puede tener diferentes significados, tales como: código, orden o cantidad. Un ejemplo en el que los numerales hacen las veces de códigos es en el caso de la representación de la variable género (nivel de medición nominal) en la cual utilizamos los números 1 (varón) y 2 (hembra). Sin embargo, los números tienen un significado de orden cuando seríamos a los estudiantes de mayor a menor o viceversa, de acuerdo con su estatura, y decimos que el más alto es el número 1, el segundo más alto es el 2, el tercero más alto, el 3 y así sucesivamente. Los números tienen un significado de cantidad cuando son utilizados en escalas de intervalos o de razón para medir los objetos, eventos o sus características; por ejemplo, la temperatura (nivel de medición de intervalo) o el peso de un objeto (nivel de medición de razón). En este sentido, un número puede ser definido como un objeto ideal al que se ha asignado un significado cuantitativo.

En tercer lugar, hay que tomar en cuenta que las reglas, que permiten asignar numerales a los objetos, pueden ser buenas o malas. En el caso de que podamos contar con “buenas reglas” tendremos “buenas mediciones”, en el sentido de lograr una aproximación al objeto que se pretende medir. Si, por el contrario, contamos con “malas reglas”, tendremos mediciones pobres o defectuosas; es decir, estimaciones muy gruesas o distorsionadas de la realidad u objeto de medición.

Muchas variables son fáciles de medir porque las reglas son sencillas de ser formuladas (por ejemplo, estado civil, tipo de vivienda, la temperatura de una persona o el peso de un objeto). Pero otras veces, como ocurre en el caso de los constructos psicológicos, las mediciones son más difíciles, porque resulta complejo establecer reglas claras y precisas que nos permitan tener una buena aproximación a la “magnitud” de nuestro objeto de interés.

## Medición y Realidad

Cuando decimos que estamos midiendo un objeto ¿cuál es realmente la estimación cuantitativa que estamos haciendo del mismo? Para algunos autores nunca medimos los objetos en sí, ni siquiera sus características o propiedades, sino más bien los indicadores de tales atributos (ver Kerlinger y Lee, 2002). Esta apreciación parece ser válida en todas las ciencias. No obstante, se destaca más en el caso particular de las mediciones de los constructos psicológicos (v.g., inteligencia, motivación, actitud, ansiedad, creatividad). Los investigadores lo que hacen es inferir estas propiedades o atributos a partir de las observaciones de supuestos indicadores de dichas características.

Otra interrogante importante en el proceso de medición se refiere a la relación necesaria que debe existir entre la realidad y su medida; es decir, ¿Hasta dónde la medida es isomórfica con respecto al objeto que pretende medir? Etimológicamente, isomorfismo significa identidad o similitud de forma. Por lo tanto, la pregunta podría ser reformulada de esta manera ¿Tiene la medida utilizada alguna correspondencia conceptual y empírica con la realidad?

En el caso de objetos físicos, la relación de isomorfismo entre la realidad y su medida es relativamente fácil de identificar, ya que el instrumento de medición actúa como un patrón de comparación en relación con la característica que se desea medir. Así, la estatura de una persona se expresa en metros o centímetros; el patrón de medida es de naturaleza lineal; no de tipo cuadrático (medida de superficie) o cúbico (medida de volumen).

La evidencia de la relación de isomorfismo entre el objeto y su medida

es más difícil de apreciar en las mediciones psicológicas, ya que nunca sabemos exactamente la verdadera naturaleza de los constructos que medimos; por ejemplo, pudiera darse el caso de que un determinado rasgo complejo que nos interesa medir es de naturaleza multidimensional; sin embargo, pudiéramos intentar medirlo con un instrumento unidimensional. Por otra parte, también pudiera darse la situación de un constructo cuya medición real se expresara en una escala del 1 al 9, pero como no lo sabemos, empleamos una escala del 1 al 5 para tal propósito. En ambos ejemplos (dimensión del constructo y tipo de escala) se presenta un problema de correspondencia entre la medida y su objeto (isomorfismo). Esta situación plantea, en el fondo, un problema de validez de la medida, la cual debe ser determinada por los medios que el investigador tenga a su alcance.

## Niveles de Medición

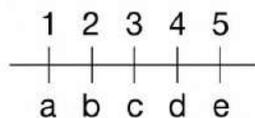
Existen cuatro niveles de medición; ellos son: nominal, ordinal, de intervalo y de razón. Estos niveles de medición conducen a igual número de escalas, en las cuales se expresan las variables que usualmente medimos en las ciencias sociales, en general, y en las ciencias de la educación, en particular.

La **medición nominal** constituye el nivel más bajo de todos los mencionados. En este caso, los objetos sólo pueden ser nombrados y contados. Consiste simplemente en clasificar observaciones dentro de ciertas categorías, las cuales deben ser mutuamente excluyentes y colectivamente exhaustivas. Por lo tanto, no puede haber ninguna observación que no pueda ser asignada a una de las categorías; y, por otra parte, una misma observación no puede ser clasificada en dos categorías diferentes, a la vez. Por ejemplo, género es una variable nominal; en consecuencia, podríamos clasificar a todos los alumnos de una sección de clase en las siguientes categorías: masculino o femenino. En tal sentido, algunos sujetos serán clasificados como masculino; mientras que otros serán ubicados en la categoría femenino, pero ninguno de los sujetos podrá ser clasificado, al mismo tiempo, en ambas categorías.

En el nivel de **medición ordinal**, las observaciones, además de poder ser clasificadas en categorías, también pueden ser ordenadas por rango, de manera creciente o decreciente. En tal sentido, una primera observación puede ser mayor que una segunda, y ésta, a su vez ser, mayor que una tercera, y así sucesivamente. Sin embargo, ello no implica una secuencia de intervalos iguales. Así, por ejemplo, podemos ordenar a tres estudiantes de acuerdo con su desempeño en una prueba de estadística; sus puntuaciones

fueron 20, 18 y 10; aquí se cumple la condición de que  $20 > 18 > 10$ , pero la distancia entre 20 y 18 no es la misma que la que existe entre 18 y 10.

El nivel de **medición de intervalo**, posee las características de los dos niveles de medición ya mencionados (nominal y ordinal), pero además contiene distancias equivalentes, que representan distancias iguales en la propiedad objeto de medición. Por ejemplo, supongamos que hemos medido cuatro objetos con una escala de intervalo y los resultados fueron 10, 8, 7 y 5. Entonces podemos decir legítimamente que la diferencia entre el primero y el tercer objeto en la propiedad medida,  $10 - 7 = 3$ , es equivalente a la diferencia entre el segundo y el cuarto objeto,  $8 - 5 = 3$ . Otra manera de expresar la idea de lo que significan los intervalos iguales consiste en decir que los intervalos pueden ser sumados y restados. Veamos el siguiente ejemplo de una escala de intervalo:



El intervalo incluido entre **a** y **c** es  $3 - 1 = 2$ . El intervalo comprendido entre **c** y **d** es  $4 - 3 = 1$ . Ahora observemos que el intervalo que va de **a** hasta **d** es  $4 - 1 = 3$ . Expresado en una ecuación:  $(d-a) = (c-a) + (d-c)$ . Si estos intervalos fueran las medidas de aprovechamiento de cinco estudiantes en una prueba de rendimiento, la diferencia entre el estudiante **a** y **c** y entre **b** y **d** deberían ser iguales; sin embargo, no podríamos decir que el rendimiento del estudiante **d** fue dos veces mayor que el del estudiante **b**.

De la misma manera, sería incorrecto decir que una persona que tiene  $40^{\circ}\text{C}$  de temperatura tiene el doble que otra que sólo tiene  $20^{\circ}\text{C}$ , o que una

persona que tiene cero (0) temperatura no tiene ninguna temperatura. En este último caso se trata de que en este nivel de medición no existe un cero absoluto, sino relativo. Por lo tanto, la medición cero no implica la ausencia de la característica estudiada.

El nivel de **medición de razón**, constituye el nivel más alto de medición. Posee las características de los niveles de medición señalados anteriormente, pero además tiene un cero absoluto que tiene un significado empírico. Si una medición con una escala de razón es igual a cero, existe razón para pensar que existen objetos que no poseen la propiedad que se pretende medir. Ahora bien, puesto que existe la presencia de un cero natural o absoluto, todas las operaciones aritméticas pueden ser ejecutadas (suma, resta, multiplicación y división). Los números de escala indican el “cuantun” de la propiedad que se pretende medir. Si existiera una escala de razón para medir el rendimiento académico, entonces sí se podría decir que un estudiante que obtuvo 20 puntos en una prueba sabe el doble de otro que sólo obtuvo 10 puntos.

## Clasificación de los Instrumentos de Medición

Los instrumentos de medición pueden ser agrupados atendiendo a diferentes criterios, así tenemos:

1. De acuerdo con el *proceso de elaboración*, pueden ser formales e informales. Los instrumentos formales son aquellos que requieren de un alto refinamiento técnico, especialmente en lo que respecta al establecimiento de normas, validez, confiabilidad y objetividad de la medida; por ejemplo, las pruebas o test, las escalas, los inventarios y los cuestionarios. Los instrumentos informales, por su parte, son aquellos cuya elaboración no requiere de tanta meticulosidad como los instrumentos formales; por ejemplo, listas de cotejo, guías de observación, guiones de entrevista, diarios de campo, entre otros.

2. Los instrumentos pueden ser agrupados también de acuerdo con su *propósito*, el cual varía de acuerdo con su tipo. Por ejemplo, las **pruebas** tienen por objeto hacer una estimación cuantitativa del comportamiento de una persona con respecto a un rasgo, atributo o característica, para lo cual los sujetos son expuestos a determinadas tareas con el propósito de provocar en ellos ciertas reacciones registrables; reacciones de toda naturaleza en cuanto a su complejidad, duración, forma, expresión y significado. Las **escalas** tienen por finalidad ubicar al sujeto en el valor escalar que mejor representa el “cuantun” del constructo objeto de medición; por lo tanto, las respuestas en este tipo de instrumento no son correctas ni incorrectas, sino que representan un valor en la escala de medición. Los **cuestionarios**, tienen como propósito

obtener información descriptiva de un conjunto de variables que permiten caracterizar una realidad social, institucional, comunal, grupal o personal.

3. De acuerdo con su *campo de aplicación*, los instrumentos pueden ser clasificados, a su vez, atendiendo a diferentes criterios; por ejemplo, tomando en cuenta su campo de aplicación, pueden ser **psicométricos** (pruebas de: inteligencia, personalidad, aptitudes), **edumétricos** (pruebas para la evaluación diagnóstica, formativa y sumativa) y **sociométricos** (pruebas de interacción social, liderazgo y cohesión del grupo).

4. Los instrumentos también pueden ser organizados de acuerdo con: (a) la forma de responder: de lápiz y papel, de manipulación, y de tipo oral; (b) el grado de objetividad en la respuesta: objetivo, subjetivo; (c) los criterios utilizados para la interpretación de sus resultados: estandarizados, no-estandarizados; (d) la forma de aplicación: individual, colectivo; y (e) el tiempo disponible para responder: test de tiempo, test de poder.

## Fases en la Construcción de un Instrumento

Como ya hemos mencionado en la sección anterior, los instrumentos de medición varían con respecto a diferentes criterios (v.g., naturaleza, propósito, formalidad). De allí que cada instrumento prácticamente requiera de procedimientos específicos para su diseño. No obstante, existen algunos procedimientos generales, al menos en los instrumentos formales, que son comunes en el diseño y desarrollo de cualquier instrumento.

En esta sección, se describe cada una de estos procedimientos generales, particularmente en lo que respecta al diseño y desarrollo de pruebas y escalas para la medición de variables psicoeducativas. Estos procedimientos son siete, a saber: (a) determinar el propósito del instrumento; (b) decidir sobre el tipo de instrumento; (c) conceptualizar el rasgo, atributo o característica (constructo); (d) operacionalizar el constructo; (e) integrar el instrumento (f) realizar el estudio piloto piloto; y (g) realizar el estudio técnico. A continuación se describe cada uno de estos procedimientos.

## Determinar el Propósito del Instrumento

El proceso de diseñar y desarrollar un instrumento de medición implica tomar un conjunto de decisiones antes, durante y después de dicho proceso. Una de las primeras y principales decisiones que, al respecto, deberá tomar el investigador se refiere a la determinación del propósito del instrumento; es decir, cuál es su finalidad, qué pretendemos medir con el instrumento; en definitiva, para qué lo queremos.

Por lo general, el propósito del instrumento está relacionado con el objetivo general del estudio o con alguno de los objetivos específicos. Por ejemplo, si el objetivo de una investigación es determinar el efecto de las estrategias metacognitivas en el rendimiento académico en matemática en estudiantes de noveno grado, uno de los posibles instrumentos a ser utilizado será una prueba para medir el rendimiento académico de los sujetos de la muestra.

En el caso anterior, el propósito o finalidad del instrumento está asociado con la necesidad de diseñar una prueba para evaluar el efecto del tratamiento (estrategias cognitivas) sobre la variable dependiente. Por supuesto, que no tiene mucho sentido ponerse a diseñar un instrumento cuando todavía no tenemos claro los objetivos de la investigación que queremos realizar.

## Decidir sobre el Tipo de Instrumento

Una segunda decisión que deberá tomar el investigador en el proceso de diseñar y desarrollar un instrumento de medición, está relacionada con el tipo de instrumento a utilizar. Como vimos en la sección anterior, existen muchas clases de instrumentos y no todos sirven para alcanzar los mismos propósitos; por lo tanto, se hace necesario pensar y decidir sobre qué instrumento utilizar y por qué. Igualmente, es importante tener claro cuáles son las ventajas y desventajas que nos ofrecen, de acuerdo con el tipo de investigación. Los alcances y limitaciones de los instrumentos que vamos a diseñar es algo que debe expresarse tanto en el proyecto como en el informe final de la investigación lo cual debemos tener muy claro antes de diseñarlos (Aroca, 1989).

En los últimos años hemos observado, con preocupación, la tendencia exagerada, de algunos estudiantes de maestría, a utilizar, en estudios descriptivos, las escalas de estimación, en lugar del cuestionario. Con el agravante de que la mayoría de dichos estudiantes tiene limitaciones para la construcción de dichas escalas y, peor aún, para el análisis e interpretación de sus resultados. Con tales procedimientos se obtienen instrumentos de dudosa validez y confiabilidad, todo lo cual conduce a resultados de investigación de poca o ninguna credibilidad.

Queda sobreentendido que en una misma investigación se pueden utilizar, no uno, sino varios procedimientos para la recolección de datos. Por ejemplo, en un estudio cuyo objetivo sea determinar la relación entre la eficiencia del docente y el rendimiento académico, mediado por el estrato socioeconómico, se podrían utilizar los siguientes instrumentos: una escala de

eficiencia docente, una prueba de rendimiento académico y un cuestionario socioeconómico.

## Conceptualizar el Constructo

Diseñar un instrumento de medición implica, entre otras cosas, tener una idea clara del objeto que se pretende medir. Para ello es indispensable hacer una revisión cuidadosa de la literatura especializada a fin de conocer si el constructo ha sido o no tratado con anterioridad. Si se tratara del primer caso, sería conveniente indagar acerca de las diferentes teorías existentes y las características de los instrumentos con que el constructo ha sido previamente medido. Asimismo, es recomendable revisar las investigaciones previas realizadas en el área. Pero, aún en este caso, pudiera ser que el investigador no se sienta satisfecho con el tratamiento que el tema ha tenido en la literatura y desea hacer una reconceptualización del constructo; en este caso, ésta debería estar precedida de una amplia y razonada justificación.

Por ejemplo, diseñar una prueba para medir habilidades cognitivas, requeriría hacer una amplia revisión de la literatura al respecto, para luego llegar a una definición teórica del constructo. Aquí se podría presentar varias posibilidades: (a) que se adopte una definición ya existente; (b) que se construya una nueva definición integrando diferentes elementos contenidos en la literatura especializada; o (c) que se formule una nueva definición del constructo totalmente original.

Si con base en la revisión de la literatura, se llegara a determinar que el constructo no ha sido previamente tratado, entonces se requiere empezar de cero, lo cual implica precisar su definición y obtener la validación respectiva, a través de juicios de expertos. En este sentido, es sumamente importante que se tome en cuenta la precisión acerca de la uni o multidimensionalidad del

objeto de medición, ya que ello constituirá la base para el desarrollo de todo el proceso de operacionalización.

## Operacionalizar el Constructo

En esta fase, es necesario transformar la conceptualización del objeto de medición en procedimientos concretos, a través de un conjunto de tareas, reactivos, preguntas o ítemes, los cuales permitirán validar empíricamente el mismo. Para tal fin, se contruye el **Plan de Operacionalización** del instrumento, el cual tiene una forma de matriz, como se ilustra con el ejemplo que se presenta en el Cuadro 1.

Como se puede apreciar en el Cuadro 1, el plan de operacionalización de un instrumento consta de cinco partes, a saber: (a) definición del propósito del instrumento; (b) definición del constructo; (c) dimensiones del constructo; (d) indicadores de cada dimensión; y (e) ítemes, reactivos, preguntas o tareas.

La definición del propósito del instrumento, plantea la necesidad de hacer explícita la finalidad del instrumento; responde a la pregunta ¿para qué queremos diseñar este instrumento? La definición del constructo se refiere a la descripción y delimitación precisa del objeto de estudio; responde a la pregunta ¿cuál es el rasgo o atributo que queremos medir?

Las dimensiones del constructo, aluden a la composición estructural que se asume teóricamente debe tener el objeto de medición; responde a la pregunta ¿es este constructo uni o multidimensional, cuáles son sus dimensiones? Los indicadores, son expresiones descriptoras de aspectos específicos relevantes que caracterizan cada una de las dimensiones del constructo; responden a la pregunta ¿cuáles son los aspectos relevantes que mejor representan esta dimensión del constructo?

*Cuadro 1. Matriz de Operacionalización de una Escala de Actitud hacia la  
Matemática*

Propósito del Instrumento	Definición del Constructo	Dimensiones	De Cada Dimensión	
			Indicadores	Ítemes (Ejemplos)
Diseñar un procesimiento sistemático para medir la actitud de los alumnos hacia el aprendizaje de la matemática.	La actitud hacia la matemática es la predisposición que tienen los alumnos para involucrarse positiva o negativamente en actividades propias de la asignatura.	1. Satisfacción con la matemática	1.1 Sentir agrado en la clase de matemática. 1.2 Disfrutar las lecturas sobre matemática. 1.3 Sentirse motivado para realizar ejercicios de matemática.	1.1.1 Me siento contento cuando estoy en la clase de matemática. 1.2.1 Disfruto mucho las lecturas sobre temas de matemática. 1.3.1 Me siento energizado cuando realizo los ejercicios de matemática con mis compañeros.
		2. Valoración de la matemática	2.1 Importancia de la matemática para el desarrollo intelectual. 2.2 Utilidad de la matemática en la comunicación entre las personas 2.3 Valor científico de la matemática	2.1.1 Considero que la practica de la matemática a contribuido con mi desarrollo intelectual. 2.2.1 El lenguaje matemático es útil en la comunicación entre los seres humanos 2.3.1 Valor científico de la matemática

*Finalmente, los ítemes, se refieren a los reactivos, tareas o preguntas a través de los cuales se operacionaliza todo el constructo y constituyen la materia prima del instrumento; responden a la pregunta ¿de que manera podemos representar operacionalmente cada uno de los indicadores de las diferentes dimensiones?*

## Integrar el Instrumento

Una vez que el constructo ha sido operacionalizado, dando lugar al conjunto de reactivos que formarán el instrumento, se procede a integrar el mismo. En este sentido, es recomendable prestar atención a: el número de ítems que serán requeridos; el tipo de reactivo (objetivo, desarrollo y mixto); la organización espacial de la información sobre el papel; la precisión de la instrucciones, la claridad en la redacción y a la extensión de la prueba. En esta fase se realiza la validez de contenido del instrumento.

## Realizar el Estudio Piloto

El proceso técnico que implica el desarrollo de un instrumento de medición puede ser realizado en corto tiempo (días, semanas) o puede durar un tiempo mayor, dependiendo del tipo de prueba y de su finalidad. Por ejemplo, la calibración de una prueba de rendimiento académico para un estudio de caso, puede realizarse en un relativo corto tiempo; mientras que una prueba de habilidades cognitivas o aptitudes, tomará mucho más tiempo.

Una vez que el constructo ha sido operacionalizado y se ha integrado la primera versión del instrumento, se pasa a la realización de la prueba piloto, la cual supone la administración del instrumento en una muestra representativa de la población donde va ser finalmente utilizada. Al respecto, es necesario hacer algunas consideraciones; primero, se requiere que el número de ítems de la prueba sea aproximadamente tres veces, en comparación con la extensión final esperada. Esto quiere decir, que si aspiramos a tener una prueba de unos 20 ítems, la primera versión de la misma debería contener unos 60 reactivos.

Segundo, se debe seleccionar una muestra de sujetos cuyo tamaño sea, como mínimo, equivalente al doble del número de ítems incluidos en el instrumento. La muestra debe ser representativa; es decir, que refleje fielmente las características de la población donde va ser aplicada (por ejemplo, edad, nivel de instrucción, sexo, estrato socioeconómico, capacidad intelectual, motivación al logro, actitudes, valores). Una manera de ganar tal representatividad en la muestra es mediante el uso de los procedimientos de muestreo estadístico.

## Estudio Técnico

Después de administrar la primera versión del instrumento se procede a realizar el estudio técnico, el cual comprende: el análisis de ítem, estimación de la confiabilidad de la medida, estudio de la validez del instrumento, estandarización y normalización.

El **análisis de ítem** normalmente se refiere al estudio de las características de los reactivos en cuanto a su nivel de facilidad, grado de discriminación y eficiencia de las alternativas de respuestas.

La **facilidad del ítem** se define como la proporción de sujetos que responden un ítem con respecto al total de sujetos que contestaron el instrumento, multiplicado por cien, como se expresa en la fórmula siguiente:

$$IF = \frac{SRCI}{N} * 100$$

Donde:

$IF$  = índice de facilidad

$SRCI$  = número de sujetos que respondieron correctamente un ítem;

$N$  = número de sujetos que respondieron la prueba.

Así, en la medida que más sujetos responden un ítem, éste tiende a ser más fácil. De la misma manera, en la medida que un ítem es respondido por pocos o ninguno de los alumnos, éste tiende a ser más difícil. El índice de dificultad de un ítem se expresa en valores que oscilan entre 0 y 100. Existen diferentes criterios para interpretar dicho índice (ver, por ejemplo, Tartarini, 1967). En nuestra opinión, una manera sencilla de interpretar el índice de dificultad de un ítem consistiría en dividir la distribución teórica del índice

(valores comprendidos entre 0 y 100) en cinco categorías con un intervalo de 20, como se indica en la escala que aparece más abajo:

81 a 100	Muy Fácil
61 a 80	Fácil
41 a 60	Moderadamente Fácil
21 a 40	Difícil
0 a 20	Muy Difícil

Veamos un ejemplo, supongamos que una prueba de rendimiento académico fue respondida por 40 estudiantes y que el ítem 1 sólo lo respondieron correctamente 10 alumnos. En este caso el índice de facilidad se calcula como sigue:

$$IF = \frac{10}{40} * 100 = 25$$

Como podemos observar el ítem fue repondido sólo por el 25 por ciento de los estudiantes, lo cual es un indicador de que dicho reactivo era difícil. La dificultad de la prueba total se puede calcular obteniendo la media aritmética de los índices de dificultad de los ítemes o utilizando la siguiente fórmula:

$$DPT = \frac{\sum ID}{Pmp}$$

Donde:

$DPT$  = Dificultad de la prueba total

$\sum ID$  = sumatoria de los índices de dificultad

$Pmp$  = Puntaje máximo posible

Por ejemplo, Suponga que en una prueba de cinco ítemes se han obtenido los siguientes índice de dificultad: 75, 25, 50, 30 y 100; la media de

la facilidad de los reactivos fue de  $M = 56$ , lo cual indica que la prueba tiene una dificultad moderada. Si se utiliza el procedimiento de la fórmula antes indicada, los datos serían los siguientes: el total de las puntuaciones de los índices de dificultad es igual a 280 y la puntuación máxima posible era de 5 puntos (un punto por cada respuesta correcta), aplicando la fórmula tendríamos:

$$DPT = \frac{280}{5} = 56$$

Como se puede apreciar, los dos procedimientos utilizados para estimar la dificultad de la prueba producen los mismos resultados.

La **discriminación**, por su parte, se refiere al poder de un reactivo para separar a aquellos sujetos que tienen dominio del rasgo que mide el ítem en comparación con aquellos que no tienen tal dominio. El índice de discriminación de un ítem es un indicador de su validez y la media de la discriminación de la prueba está directamente relacionada con su confiabilidad. Mientras mayor sea este valor promedio, mayor será la confiabilidad de la prueba.

Existen diferentes procedimientos para calcular el índice de discriminación de los ítems de un instrumento de medición. Por ejemplo, si se tratase de una prueba de rendimiento académico, de habilidad cognitiva (inteligencia general, habilidades específicas, aptitudes) u otras, cuyos ítems sólo presenten dos opciones de respuesta posibles (correcto versus incorrecto), el procedimiento sería el indicado por Dawson y Thomas (1972), el cual se expresa en la fórmula siguiente:

$$ID = \frac{RCGS - RCGI}{N_{gi}}$$

Donde:

$ID$  = índice de discriminación del ítem;

$RCGS$  = número de respuestas correctas en el grupo superior;

$RCGS$  = número de respuestas correctas en el grupo inferior;

$N_{gi}$  = tamaño de uno de los grupos.

Para comprender mejor lo que significa la discriminación de un ítem vamos a desarrollar un ejemplo. Supongamos que se ha administrado una prueba de rendimiento académico (6 ítems) a un grupo de  $n = 12$  estudiantes, cuyos resultados se presentan en el Cuadro 2.

*Cuadro 2. Matriz de ítem por sujeto de los resultados de una prueba de rendimiento académico*

Sujetos	Ítems						Total
	1	2	3	4	5	6	
Pérez, J.	✓	✓	✓	✓	✓	✓	6
García, M.	✓	✓	✓	✓	✓	✓	6
Ruiz, P.	✓	x	✓	✓	✓	✓	5
Hernández, H.	x	✓	✓	✓	✓	✓	5
Bolívar, C.	✓	x	✓	✓	x	✓	4
Sifontes, R.	x	x	✓	✓	✓	✓	4
Grance, B.	✓	✓	✓	x	x	x	3
Manrique, W.	x	x	x	✓	✓	✓	3
Rivero, G.	✓	✓	x	x	x	x	2
Rendón, P.	x	x	✓	✓	x	x	2
Gonzalez, M.	x	x	x	x	✓	✓	2
Echenique, A.	✓	x	x	x	x	x	1

x = incorrecta    ✓ = correcta

Observe que la matriz de datos está ordenada jerárquicamente, de acuerdo con el valor de las calificaciones, y que la misma ha sido dividida en tres subgrupos de alumnos, de acuerdo con los resultados obtenidos por ellos en la prueba. El subgrupo que obtuvo puntajes comprendidos entre 5 y 6, se

denomina grupo superior (GS); mientras que quienes obtuvieron puntajes comprendidos entre 1 y 2 se denominan grupo inferior (GI).

Supongamos que queremos estimar el índice de discriminación del ítem 6; en este caso procederíamos de la siguiente manera:

$$ID = \frac{4-1}{4} = 0,75$$

Observe que el ítem 6 fue respondido correctamente por todos los alumnos ( $n = 4$ ) del grupo superior; mientras que sólo fue acertado por un alumno en el grupo inferior. Al aplicar la fórmula correspondiente obtenemos un índice de 0,75. ¿Cómo se interpreta este valor? En general, el índice de discriminación de un ítem puede ser interpretado atendiendo a la siguiente escala:

0,81 a 1,00	Muy Alto
0,61 a 0,80	Alto
0,41 a 0,60	Moderado
0,21 a 0,40	Bajo
0 a 0,20	Muy Bajo

Con base en la escala anterior, el valor 0,75 indica que el ítem 6 tiene un poder discriminativo alto. El valor del ID oscila entre +1 (discriminación perfecta) y -1 (discriminación negativa). Un ítem tiene una discriminación perfecta cuando el mismo es respondido por todos los sujetos del grupo superior y por ninguno de los del grupo inferior. Cuando esta relación es inversa; es decir, que el ítem no lo pasa nadie en el grupo superior y lo responden correctamente todos los sujetos del grupo inferior, el ítem tiene una discriminación negativa y, en consecuencia, debe ser descartado.

Cuando un ítem es respondido correctamente por todos los sujetos en ambos grupos, su discriminación es igual a cero.

Otra manera de calcular el ID consiste en transformar el número de respuestas correctas en ambos grupos (superior e inferior) a porcentajes y luego calcular la diferencia entre ellos. En el caso del ítem 6, al que ya hemos hecho referencia en el Cuadro 2, tendríamos: en el grupo superior  $4/4 * 100=100$  y en el grupo inferior  $1/4 * 100=25$ ; luego,  $100 - 25 = 75$ . Como se puede apreciar este resultado es idéntico al anterior, sólo que el primero fue expresado como una proporción o razón; mientras que este último se calculó en términos porcentuales (Anastasi, 1961).

Para interpretar el ID, de acuerdo con este segundo procedimiento, los valores de la escala se multiplican por 100; así un ID será muy alto, cuando se encuentre comprendido entre los valores 81 y 100 por ciento. Por otra parte, cuando se trata de instrumentos, en los cuales no existe la posibilidad de respuestas binarias o dicotómicas, como es el caso de las escalas de actitud, el procedimiento para calcular la discriminación de los ítems es mediante el uso de la prueba t de Student para muestras independientes, como se indica en el Capítulo 3 de esta obra.

La facilidad (o dificultad) y la discriminación de los ítems de un instrumento de medición son dos características que no son independientes, sino que están íntimamente relacionadas. En tal sentido, de acuerdo con las definiciones formuladas en cada caso, si un ítem de una prueba lo pasa el 100 por ciento de los sujetos, es un reactivo muy fácil, pero al mismo tiempo su discriminación es cero (0). Asimismo, si un ítem no lo pasa ninguno de los sujetos que tomaron la prueba, significa que el reactivo es muy difícil, pero su discriminación, también, es cero (0). Es por ello que estas dos características técnicas de los ítems se utilizan de manera combinada en el

proceso de selección de los reactivos que integrarán la versión final del instrumento.

Algunos de los criterios que corrientemente se utilizan en la toma de decisiones, en relación con la facilidad y discriminación de los ítemes, son los siguientes:

1. Seleccionar ítemes que sean moderadamente fáciles (entre 41 y 60 por ciento) y que al mismo tiempo, tengan una discriminación moderada o muy alta (índices entre 0,41 y 1.0).
2. Seleccionar ítemes con índices de facilidad y discriminación moderados (índices entre 41 y 60 por ciento).
3. Seleccionar ítemes con índices de facilidad alto, moderado y bajo, manteniendo un nivel de discriminación entre bajo y moderado.
4. Seleccionar ítemes con índices de facilidad baja (reactivos difíciles) y que, al mismo tiempo, tienen un nivel de discriminación moderado.

En relación con estos criterios, es importante señalar que ninguno de ellos es mejor o peor que otro; la decisión acerca de cuál de ellos se debe utilizar en momento dado, depende del tipo de instrumento y del propósito del investigador. Por ejemplo, en las pruebas corrientes utilizadas para evaluar el rendimiento académico, es bastante común utilizar el criterio 1. Pero, en el caso de pruebas de aptitudes o de habilidades, cuando se utilizan con propósitos de selección, por lo general se prefiere el criterio 4.

Para ilustrar cómo se procede para seleccionar los ítemes de acuerdo con sus niveles de facilidad y discriminación, se presenta un ejemplo numérico. Supongamos que hemos administrado una prueba de 10 reactivos en una muestra de  $n = 30$  estudiantes y que los resultados del análisis de ítem se presenta en el Cuadro 3.

En el Cuadro 3 se puede observar que hay seis ítemes con índices de facilidad moderado y seis reactivos con índices de discriminación ubicados entre moderados y altos. Ahora bien, si seleccionáramos estos ítemes con base en el criterio 1, ya mencionados, sólo los reactivos N° 4, 6, 7 y 9 serían incluidos en la versión final de la prueba, ya que son los únicos que satisfacen el criterio de tener un índice de de facilidad moderada y una discriminación entre moderada y alta.

*Cuadro 3. Indices de facilidad y discriminación para una prueba de rendimiento académico en Castellano (n = 30)*

Item	Indice de Facilidad		Indice de Discriminació	
1	100		0,00	
2	80		0,45	*
3	58	*	0,38	
4	45	*	0,76	*
5	10		0,80	*
6	60	*	0,45	*
7	55	*	0,43	*
8	57	*	0,31	
9	48	*	0,52	*
10	0		0,00	

(\*) satisface el criterio

Pero, por otra parte, para que los ítemes de un instrumento de medición, sean considerados como “buenos”, en el sentido de que pueden ser incluidos en la versión final de la prueba, no es suficiente con que tengan los índices de facilidad y discriminación apropiados, sino que se requiere, además, que las **alternativas de respuestas funcionen de manera balanceada**; es decir, que todas las alternativas de respuestas estén incluidas en las distribuciones de los reactivos. Teóricamente se espera que las respuestas dadas a un ítem deberían distribuirse más o menos proporcionalmente a través de las diferentes

alternativas de respuestas del reactivo. Por ejemplo, si una prueba objetiva de selección simple, con cuatro opciones de respuesta por ítem, fue tomada por 35 estudiantes, de los cuales sólo cinco respondieron correctamente el reactivo 1, es de esperarse que las respuestas de los 30 estudiantes restantes se distribuyan de manera balanceada a través de los diferentes distractores del ítem, como se indica en el Cuadro 4.

*Cuadro 4. Ejemplo de posibles distribuciones aceptables de respuesta dadas a un ítem*

Alternativas de Respuesta	Posibles Distribuciones de las Respuestas			
	1	2	3	4
a*	5	5	5	5
b	10	8	5	20
c	10	12	15	5
d	10	10	10	5
Total	35	35	35	35

(\*) Respuesta Correcta

Observe que de todas las cuatro distribuciones, presentadas en el Cuadro 4, la N° 1 es teóricamente la mejor; mientras que la N° 4 parecería la peor. Sin embargo, en la práctica una distribución perfecta como la N° 1, en la cual las respuestas incorrectas se distribuyeron por igual en los tres distractores, es bastante improbable. Pero no se trata de lograr siempre este tipo de distribución; más bien lo que se espera es que cada distractor capte, por lo menos, entre un 10 y un 15 por ciento de las respuestas incorrectas. Por lo tanto, nunca se espera lograr distribuciones como las que se indican en el Cuadro 5.

*Cuadro 5. Ejemplo de Posibles Distribuciones de Respuestas No Aceptables*

### *Dadas a un Item*

Alternativas de Respuesta	Posibles Distribuciones de las Respuestas			
	1	2	3	4
a*	5	5	5	5
b	30	0	28	3
c	0	5	0	27
d	0	25	2	0
Total	35	35	35	35

(\*) Respuesta Correcta

Note que en el Cuadro 5, las alternativas de respuestas incorrectas tienden a estar concentradas en uno o dos distractores y no en los tres, como debería ser. Este tipo de situación debe ser evitada durante el proceso de calibración de un instrumento de medición, ya que ello indica que el hecho de que algunos distractores no capten ninguna respuesta se debe a que los sujetos lo pueden identificar como tal y, en consecuencia, lo evitan. Ello se interpreta como una deficiencia en el funcionamiento del ítem, por lo cual debe ser sometido a revisión, mejoramiento y reanálisis hasta lograr su adecuación a los requerimientos técnicos del instrumento.

## Estandarización

Es normal en el proceso de medición que los resultados obtenidos con un instrumento puedan ser utilizados para comparar el desempeño de una persona consigo misma, a través del tiempo, o con otras personas en un momento determinado. Pero, para que ello tenga sentido, es necesario controlar las posibles fuentes de errores que potencialmente pueden contaminar dichos resultados. Esto se logra, haciendo que la situación de prueba sea lo más semejante posible para todos los individuos. El proceso de desarrollo de tales controles se denomina **estandarización**, la cual está sujeta a tres condiciones básicas, que son:

1. Que a cada sujeto le sean presentados los mismos reactivos (o equivalentes); es lo que se denomina control de contenido.
2. Que existan reglas específicas para la administración y calificación de los ítems; y
3. Que se establezcan normas para la aplicación generalizada del instrumento.

### **Control del Contenido**

A menos que se pruebe a todos los sujetos utilizando ítems iguales o equivalentes, no será posible comparar directamente su ejecución, ya que los resultados se basarán en diferentes muestras de reactivos (Brown, 1980).

### **Condiciones de Administración y Calificación**

Aun cuando se esté utilizando la misma muestra de reactivos con todos los sujetos, las calificaciones no podrían ser comparables a menos que se aplicaran bajo las mismas condiciones.

Por supuesto, es imposible mantener un control absoluto de todos los

factores que pueden inducir a error en la medición, debido a que una prueba se aplicará a diferentes personas, en diversos tiempos y lugares y por diferentes examinadores. No obstante, es posible eliminar gran parte de la varianza de error, utilizando instrucciones prescritas para la aplicación de las pruebas, límite de tiempo estándar y métodos objetivos de calificación.

En relación con el uso de las instrucciones claras y precisas en los instrumentos de medición, por lo general, éstas son de dos tipos: (a) uno dirigido a la persona que responde la prueba; y (b) otro dirigido al administrador del instrumento. En el primer caso se debe explicar, de la manera más simple y sencilla el modo como el sujeto deberá responder los reactivos. Estas instrucciones se escriben al comienzo de la prueba, preferiblemente en una página separada. Corrientemente, estas instrucciones incluyen, además de las indicaciones para responder los reactivos, información sobre el límite de tiempo e ítemes de práctica.

A continuación se presentan algunos ejemplos de tipos de instrucciones que se utilizan en diferentes instrumentos de medición psicoeducativa.

### *Ejemplo 1*

<p><b>Instrucciones para una Escala de Actitud hacia la Lectura</b></p> <p>Lea cada uno de los ítemes que aparecen más abajo y marque con una equis (x) la alternativa de respuesta que mejor represente su situación.</p> <p>Encuentro gran satisfacción en la lectura</p> <p>Si <input type="checkbox"/>      No <input type="checkbox"/></p> <p style="text-align: right;"><i>(Ruiz, 1994)</i></p>
---

### *Ejemplo 2*

### Instrucciones para un Test de Inteligencia

Este cuadernillo tiene cuatro pruebas, que les proporcionarán a ustedes la posibilidad de demostrar lo que saben y como razonan. Ustedes tiene que responder a todas las preguntas de este cuadernillo marcando con una equis (x) la alternativa de respuesta que considere correcta. Cada pregunta va seguida de cinco respuestas, una sola de cuales es la verdadera

Fíjense en la primera de las preguntas, cuya respuesta verdadera ha sido bien señalada. Estúdienla con atención, para que comprendan cómo deben responder las demás preguntas.

1. Escoja la palabra que tiene el mismo o casi el mismo significado que la palabra que figura en caracteres gruesos en el margen izquierdo de la página:

**PERRO**

- |              |                                     |
|--------------|-------------------------------------|
| A aterrizado | <input type="checkbox"/>            |
| B canción    | <input type="checkbox"/>            |
| C animal     | <input checked="" type="checkbox"/> |
| D grande     | <input type="checkbox"/>            |
| E mosca      | <input type="checkbox"/>            |

*(Lorge y Thorndike, 1975)*

Las instrucciones para quien administra la prueba se escriben en un Manual por separado. Estas indicaciones incluyen: las que se dan al sujeto, pero con una información amplia acerca de sus implicaciones, además se incluyen recomendaciones acerca de la disposición del ambiente de aplicación, la distribución del material, el registro del tiempo y de las calificaciones. También se dan instrucciones sobre cómo manejar los problemas y la preguntas que se presenten durante la administración del instrumento (ver Ejemplo 3).

*Ejemplo 3*

### **Instrucciones para el Administrador**

Esta prueba debe administrarse en un sitio con adecuadas condiciones de luz, ventilación, temperatura y silencio. Los examinandos deben colocarse en sillas suficientemente cómodas. Conviene que el cuestionario y la hoja de respuesta se puedan colocar, de preferencia, sobre una mesa. Una vez ubicados los examinandos en el sitio que le corresponde para el examen, se les entrega la hoja de respuesta y se les solicita que escriban solamente en la parte de los datos personales. El examinador cuidará que los datos de esta parte sean correctamente llenados.

- 1.** El cuestionario tiene tres partes, cada una de ella está precedida de un conjunto de instrucciones que deben leerse con atención. No debe comenzarse a responder hasta tener una comprensión suficiente de las instrucciones. En caso de duda debe consultarse al examinador.
- 2.** Por ninguna razón deben escribirse marcas o respuestas en el cuadernillo.
- 3.** Tampoco debe marcarse las respuestas de los ejemplos en la hoja de respuesta, ni en el cuadernillo.
- 4.** El tiempo disponible para trabajar en el cuestionario es de 20 a 35 minutos.
- 5.** Cada examinando leerá las instrucciones individualmente. En caso de que se le solicite, examinador dará instrucciones que complementen lo escrito.
- 6.** El examinador estará atento a responder preguntas, evitando sugerir respuestas al examinado.
- 7.** El examinador igualmente estará atento para evitar el deterioro de los cuadernillos con marcas o respuestas.

Además, se observará que las personas marquen correctamente en la hoja de respuesta.

Para la calificación se utilizan tres rejillas, cada de las cuales recoge datos de cinco factores. El procedimiento consiste en colocar cada rejilla sobre la hoja de respuesta, sumar los valores que aparecen en cada una de las variables y anotarlos en la columna de la derecha de la hoja de respuesta. Para elaborar el perfil se convierte el puntaje bruto (PB) en puntaje estándar (PT), consultando la tabla correspondiente.

(Toro, 1992)

Otro aspecto relacionado con las condiciones de aplicación de un instrumento de medición se refiere al *control del tiempo*. En las mediciones psicoeducativas en control de tiempo puede ser usado de dos maneras: (a) en

el caso de las pruebas de **poder**, en el cual el tiempo es libre; es decir, al sujeto se le da un tiempo ilimitado para responder los reactivos presentados en orden de dificultad creciente; y (b) en el caso de las pruebas de tiempo, como las que miden aptitudes o rendimiento académico, en las cuales la velocidad es considerada como un factor importante. En muchos casos, el límite de tiempo que se establece obedece a razones de tipo prácticas, tales como la duración de los períodos de tiempo de una clase. Por regla general, en los casos de las pruebas de rendimiento académico y aptitudes se establecen límites de tiempo que permitan que aproximadamente el 90 por ciento de los sujetos concluyan en el tiempo previsto. Para establecer el criterio de tiempo en el cual debe ser administrada una prueba es importante que durante la administración experimental (estudio piloto) de la prueba se registre el tiempo que cada sujeto tardó para responder el instrumento. Luego con esta información se puede obtener el tiempo promedio en que, por lo menos, el 90 por ciento de la muestra de sujetos tarda para responder la prueba.

### **Calificaciones Objetivas**

En este contexto, la objetividad se refiere al máximo acuerdo que debe existir entre diferentes calificadores de un mismo instrumento. Idealmente, este acuerdo debería ser perfecto. No obstante, en la práctica, sobre todo en el caso de los reactivos de respuestas libres, dicho acuerdo es siempre menos que perfecto. Una regla empírica consiste en considerar la calificación como objetiva si existe, por lo menos, un 90 por ciento de acuerdo entre los calificadores. Sólo en el caso de que la calificación sea objetiva se podrán atribuir las diferencias entre calificaciones a las diferencias existentes entre los sujetos examinados.

De acuerdo con Brown (1980), las exigencias para la calificación

objetiva de una prueba son de tres tipos.

1. Registro inmediato y carente de ambigüedades, de la respuesta. Tanto si el sujeto efectúa una marca en una hoja de respuesta, como si escribe una letra, una palabra, un número, una frase o un ensayo, o responde en forma oral, su respuesta se registrará de manera inmediata y completa. Este registro permanente evitará las distorsiones posibles debidas a pérdidas de la memoria y proporciona las bases para la calificación de las respuestas.

2. Utilización de una clave de respuesta. Para un examen de selección múltiple esta clave tiene que incluir el número o la letra de la respuesta correcta a cada pregunta; para los reactivos de respuesta cortas se tratará de una lista de respuestas correctas y las variaciones aceptables; para un ensayo, se da un bosquejo de los puntos que se deben tratar. En el caso de un inventario de la personalidad, la clave designará las respuestas que indiquen la presencia (o ausencia) del rasgo o la característica que se mida. Si respuestas diferentes reciben pesos distintos, tales pesos deben indicarse en la clave.

3. Utilización de un procedimiento para comparar las respuestas dadas por el sujeto con las que figuran en la clave, esto es un procedimiento para clasificar objetivamente las respuestas. En los ítemes de elección alterna (v.g., selección múltiple), este procedimiento es directo y evidente. Cuando el buen juicio del calificador puede ser uno de los factores importantes, como sucede al calificar exámenes de ensayo, es preciso desarrollar indicaciones detalladas para la asignación de las calificaciones.

### **Establecimiento de Normas**

Las normas son criterios preestablecidos que se utilizan como punto de

comparación de las puntuaciones individuales que obtienen los sujetos en las pruebas, con respecto al desempeño de un grupo representativo que se toma como referencia (grupo normativo).

Los resultados de las pruebas tienen poco significado por sí mismos. Ellos adquieren significado cuando se compara una calificación individual con los resultados obtenidos por otras personas en la misma prueba. Este proceso de comparación nos permite obtener una indicación de su desempeño relativo en comparación con el de otros de la misma población. De allí que sea necesario establecer, como parte del proceso de desarrollo del instrumento, el conjunto de normas que permitirán interpretar los resultados en sus aplicaciones futuras.

En otras palabras, el grupo normativo proporciona una base de comparación que muestra las calificaciones de un grupo estándar, definido, de referencia. Teóricamente existen para cada prueba distintos posibles grupos normativos, lo cual hace que la clasificación relativa de un sujeto varíe mucho, dependiendo del grupo normativo que haya sido utilizado para la comparación.

Por lo tanto, la composición del grupo normativo es de crucial importancia para la interpretación de las calificaciones relacionadas con las normas. De allí que, previo al desarrollo de un conjunto de normas, lo primero que habría que definir es la población donde será aplicado el instrumento. Los grupos normativos se deben elegir para representar esas poblaciones. Por ejemplo, si el instrumento ha sido diseñado para evaluar la aptitud académica de los estudiantes de educación media que van a ingresar al subsistema de educación superior, el grupo normativo debe ser seleccionado de la población de estudiantes de educación media. Para las

pruebas de aptitudes y rendimiento académico, el grupo normativo apropiado consiste, generalmente, en los competidores existentes y potenciales.

En el caso de las pruebas de habilidades generales y de personalidad, los grupos normativos se componen habitualmente de personas de la misma edad o el mismo nivel educativo. Otras variables que usualmente son consideradas para establecer las normas de una prueba son: el sexo, la ocupación, la zona geográfica y la posición socioeconómica.

Algunos de los requisitos para seleccionar un grupo normativo son los siguientes:

### **Primero**

La composición de las normas de grupo debe ser definida con claridad. Aun cuando las especificaciones generales de los grupos normativos serán determinados en función de la finalidad y uso de la prueba, dentro de ese rango existen potencialmente diferentes grupos normativos, por ello es que se requiere de una descripción precisa y clara de la naturaleza y las características de cada grupo normativo.

Por ejemplo, no sería suficiente decir, en el caso de la prueba de aptitud académica (PAA), que el grupo normativo estuvo formado por una muestra de estudiantes de educación media de una población de 80.000 que se preinscribió para tomar dicha prueba en el año de 1980.

Habría que decir que: el grupo normativo estuvo constituido por una muestra, estratificada, con afijación proporcional, de  $n = 10.000$  estudiantes de quinto año de educación media, varones y hembras, de colegios públicos y privados, de todas las regiones del país, que tomaron la PAA durante el primer semestre del año de 1980.

### **Segundo**

El grupo normativo debe constituir una muestra representativa de la

población designada, en la cual los procedimientos de muestreo queden claramente establecidos.

### **Tercero**

Las normas deben ser actualizadas periódicamente. Es decir, las normas pueden perder vigencia, debido a cambios rápidos que se producen en la sociedad, en general, y en particular en el mundo de la educación y del trabajo.

Existen varios tipos de calificaciones normativas, como son: percentiles, calificaciones estándar, escalas de desarrollo, índices y cocientes. En esta sección sólo nos referiremos a los dos primeros tipos de calificaciones normativas, por ser los más usados en la mayoría de los casos.

El **rango percentil** de una calificación se define como el porcentaje de casos que se encuentran por debajo de dicha calificación; es decir, es un tipo de medición ordinal, ya que permite jerarquizar a los sujetos de acuerdo con su posición relativa obtenida en el grupo normativo de referencia, con base en la calificación recibida. Por ejemplo, cuando decimos que el puntaje de un alumno, en una prueba de rendimiento académico, fue de 15 puntos y ésta corresponde al percentil 50; lo que queremos decir es que el 50 por ciento del grupo obtuvo calificaciones iguales y por debajo de 15 puntos.

En el caso de las normas percentiles de la ESCAHINE (Ruiz, 1991), la tabla de normas indica que las puntuaciones 96, 102 y 128, corresponden a los percentiles 25, 50 y 75 respectivamente. De tal manera que si esta escala se aplicara a una persona cuya puntuación en la misma fuera, por ejemplo, de 130, tal resultado indica que esta persona tuvo un desempeño correspondiente al 25 por ciento superior con respecto al grupo normativo.

Las **calificaciones estándares**, son mediciones intervalares basadas en transformaciones lineales de los puntajes brutos o directos obtenidos en una

medida. La transformación lineal implica cambiar la escala de manera que se modifique la media y la desviación estándar, mientras se conserva exactamente la forma de la distribución y, en consecuencia, las posiciones relativas de los individuos en dicha distribución (Magnusson, 1982). Existen varios tipos de escalas estandarizadas, entre las cuales se encuentran, z, T (o Z), estaninas y C.I equivalentes. En esta sección sólo nos referiremos a las dos primeras.

La **escala z**, permite obtener puntuaciones estandarizadas, cuyos valores, en un 99,9 por ciento de los casos oscilan entre más o menos 3. Las puntuaciones z se obtienen restando cada puntaje individual de la media de la distribución y dividiendo, tal resultado, entre la desviación estándar, como se indica en la ecuación siguiente:

$$z = \frac{X-M}{s}$$

Donde:

$z$  = puntaje estandarizado

$X$  = puntaje bruto individual

$M$  = media de la distribución; y

$s$  = desviación estándar de la distribución

Por ejemplo, un alumno que obtuvo una calificación de 12 puntos en una sección de clase donde la media fue de 15 y la desviación estándar fue igual a 2, tiene un puntaje estandarizado ( $z$ ) igual a -1,5, de acuerdo con la fórmula anterior, como se indica más abajo:

$$z = \frac{12-15}{2} = \frac{-3}{2} = -1,5$$

Lo anterior significa que este alumno está ubicado una y media (1,5) desviación estándar por debajo de la media de la distribución. Al respecto,v

es importante señalar que puesto que en la distribución  $z$ , la media es igual a cero, los valores negativos sólo indican que un puntaje dado está ubicado por debajo de la media del grupo; mientras que un puntaje con signo positivo indicaría que la calificación está por encima de la media.

La **escala T**, permite transformar las puntuaciones originales (puntajes brutos) a una distribución cuya media es 50 y su desviación estándar es igual a 10. Sus valores, en un 99,9 por ciento de los casos caen entre las puntuaciones 20 y 80. Su fórmula es la siguiente:

$$T = 50 + 10z$$

En el caso anterior, una calificación de 12 puntos (con un valor de  $z = -1,5$ ) es igual a 35, como se indica más abajo:

$$T = 50 + 10(-1,5) = 50 - 15 = 35$$

La interpretación de las puntuaciones de T es equivalente a la indicada en el caso de  $z$ . No obstante, para muchos autores le resulta más cómodo trabajar con la escala T, puesto que no tiene valores negativos, como si ocurre en el modelo de  $z$ .

En el caso de la normalización de la ESCAHINE, los puntajes originales de la escala oscilaron entre 87 y 146 puntos, con una media de 121. Estos valores, al ser transformados a través de la escala de T, correspondieron 25,71; 67,87 y 50, respectivamente (ver Ruiz, 1991).

# Confiabilidad y Validez

Al evaluar la utilidad de una prueba o test, a menudo hay dos preguntas que es necesario formularse, las cuales aunque son diferentes, de alguna manera están relacionadas. La primera pregunta es: ¿con cuánta exactitud la muestra de ítems o tareas representa al universo de donde fueron seleccionados? La segunda pregunta es ¿con qué fidelidad corresponde este universo al atributo latente que se va a medir? La primera pregunta se relaciona con lo que comúnmente se denomina **confiabilidad** de la medida; mientras que la segunda se refiere a su **validez**.

En este capítulo nos referiremos, particularmente, a estas dos características de las pruebas y escalas, por ser éstos los instrumentos que plantean mayores exigencias técnicas en su proceso de desarrollo. Estas dos características son fundamentales cuando se trata de medir rasgos o atributos psicológicos. De allí que deban ser cuidadosamente establecidas en las llamadas pruebas formales, como parte de lo que hemos denominado el estudio técnico del instrumento.

## Confiabilidad

Una de las características técnicas que determinan la utilidad de los resultados de un instrumento de medición es su grado de **reproducibilidad**. Esta se refiere al hecho de que los resultados obtenidos con el instrumento en una determinada ocasión, bajo ciertas condiciones, deberían ser similares si volviéramos a medir el mismo rasgo en condiciones idénticas.

Este aspecto de la **exactitud** con que un instrumento mide lo que se pretende medir es lo que se denomina la **confiabilidad** de la medida. En este sentido, el término confiabilidad es equivalente a los de **estabilidad** y **predictibilidad**. Esta es la acepción que más comúnmente se le da a este término.

No obstante, otra manera de aproximarse al concepto de confiabilidad es preguntarse: ¿Hasta dónde los resultados obtenidos con un instrumento de medición constituyen la medida “verdadera” de la propiedad que se pretende medir? Esta acepción del término confiabilidad es sinónimo de **seguridad**.

Todavía existe una tercera posibilidad de cómo podemos enfocar la confiabilidad de un instrumento de medición; ella responde a la siguiente cuestión: ¿cuánto error está implícito en la medición de un instrumento? Se entiende que un instrumento es menos confiable en la medida que hay un mayor margen de error implícito en la medición. De acuerdo con esto, la confiabilidad puede ser definida como la ausencia relativa de error de medición en el instrumento; es decir, en este contexto, el término confiabilidad es sinónimo de **precisión**.

Finalmente, la confiabilidad puede ser enfocada como el grado de homogeneidad de los ítemes del instrumento en relación con la característica

que pretende medir. Es lo que se denomina la confiabilidad de **consistencia interna** u **homogeneidad**.

### **El Modelo Clásico de Confiabilidad**

De acuerdo con la teoría clásica, toda medición que se hace de un objeto físico o psicológico lleva implícito un error; es decir, se considera que toda puntuación que se obtiene en una prueba está formada por dos componentes aditivos, que son la puntuación “verdadera” y el “error” aleatorio.

La puntuación verdadera se define como “el valor promedio de un número grande de administración de una prueba a la misma persona (Thorndike, 1989; Kerlinger y Lee 2002). Es decir, la puntuación verdadera puede ser considerada como la puntuación obtenida por un individuo si todas las condiciones, tanto internas como externas, estuvieran controladas y el instrumento de medición fuera “perfecto”. Pero, esta condición es sólo hipotética, ya que como sabemos en toda medición hay un error implícito.

La definición anterior tiene un valor meramente teórico, ya que en la práctica nunca se puede conocer directamente el componente de medición verdadera de la característica de un objeto. El error, por su parte, se refiere al aumento o disminución de la medición, como resultado de diferentes factores que determinan el error de medición; éstos dependen, algunas veces, de una validación inadecuada del instrumento de medida y otras, de las condiciones externas bajo las cuales se realiza la medición. Esta relación puede ser expresada en los términos siguientes:

$$X_t = X_v - X_e$$

Donde:

$X_t$ =puntaje total

$X_v$  = puntaje verdadero

$X_e$  = puntaje de error

Esta ecuación permite estimar el componente de puntuación verdadera en una medición, el cual estaría determinado por la diferencia entre el puntaje total obtenido y la puntuación de error; o sea,

$$V_v = V_t - V_e$$

En término de varianza, los planteamientos anteriores pueden ser expresados como sigue: la varianza total ( $V_t$ ) de una distribución de mediciones se descompone en una varianza verdadera ( $V_v$ ) y una varianza de error ( $V_e$ ), lo cual nos llevaría a la fórmula siguiente:

$$V_t = V_v + V_e$$

En consecuencia,

$$V_v = V_t - V_e$$

En resumen, la confiabilidad enfocada desde el punto de vista de la teoría del error de medición, nos llevaría a establecer una relación inversa con respecto a la confiabilidad, en los términos siguientes: a mayor error implícito en la medición, menor confiabilidad; mientras que a menor error, mayor confiabilidad. En términos prácticos, esto significa que si podemos estimar la varianza de error de una medida, también podemos estimar su confiabilidad. Todo lo cual nos lleva a una reconceptualización de la confiabilidad en el sentido siguiente:

1. La confiabilidad puede ser vista como la proporción de la varianza “verdadera” con respecto a la varianza total obtenida en la administración

de un instrumento de medición, o sea:

$$r_{tt} = \frac{V_v}{V_t}$$

2. La confiabilidad puede ser entendida como la diferencia entre 1.00 y la proporción del error de la varianza con respecto a la varianza total obtenida en la administración de un instrumento de medición, o sea:

$$r_{tt} = 1 - \frac{V_e}{V_t}$$

Como el lector especializado debe recordar, el modelo de análisis de varianza (ANAVA) de una vía permite descomponer la varianza total de la variable dependiente en dos tipos de varianzas: la varianza entre grupos (varianza experimental o verdadera) y la varianza de error. En consecuencia, este modelo de ANAVA podría ser utilizado para estimar:  $V_t$ ,  $V_v$  y  $V_e$ , y de esa manera calcular el coeficiente de confiabilidad de un instrumento de medición (ver Hoyt, 1941 y Kerlinger y Lee (2002)).

Otra manera de estimar la confiabilidad, a partir de la varianza de error, es estimando directamente el error estándar de medición, como sigue:

$$E_{EM} = S_y \sqrt{1 - r_{tt}}$$

Donde:

$E_{EM}$  = error estándar de medición

$S_y$  = desviación estándar de la distribución de puntajes totales de la prueba

$r_{tt}$  = coeficiente de confiabilidad

### **Procedimientos para Estimar la Confiabilidad**

Existen varias maneras para estimar la confiabilidad de una medida. En esta sección desarrollaremos tres de las más conocidas: (a) confiabilidad de **reaplicación de pruebas** (test-retest); (b) confiabilidad de **versiones**

**equivalentes** (pruebas paralelas); y (c) confiabilidad de **consistencia interna** (homogeneidad).

**La confiabilidad de reaplicación de pruebas.** Esta consiste en administrar dos veces una misma prueba a un mismo grupo de sujetos en un intervalo relativamente corto de tiempo (no más de tres meses entre una y otra medición). Estas dos distribuciones de puntajes se correlacionan y el coeficiente obtenido representa una estimación de la confiabilidad del instrumento (ver Anastasi, 1976).

En el caso anterior, el error de la varianza corresponde a las fluctuaciones aleatorias en las puntuaciones obtenidas entre la primera y la segunda aplicación de la prueba. Estas variaciones resultan, en parte, de ciertas condiciones no controladas durante la administración de la prueba, como, por ejemplo, ruidos, cambios de temperatura o cualquier otro distractor. También, estas variaciones pueden ser debidas a cambios ocurridos en los propios sujetos, tales como: enfermedades, fatiga, situación emocional, preocupación, estrés, etc).

La confiabilidad de reaplicación de pruebas muestra hasta donde los puntajes obtenidos en un instrumento pueden ser generalizados a través del tiempo. En la medida que la confiabilidad es mayor, menos susceptibles son los puntajes de ser modificados por las condiciones aleatorias asociadas con la situación de medición o con los cambios de los propios sujetos. El coeficiente de confiabilidad obtenido es una medida de la estabilidad de la prueba. A continuación se presenta un ejemplo acerca del cálculo de este tipo de coeficiente de confiabilidad (ver Cuadro 6).

*Cuadro 6. Estimación del coeficiente de confiabilidad de una prueba por el método de reaplicación (test-retest)*

Sujetos	1ª Aplicación		2ª Aplicación		XY
	X	X <sup>2</sup>	Y	Y <sup>2</sup>	
1	35	1225	36	1296	1260
2	40	1600	38	1444	1520
3	20	400	19	361	380
4	30	900	30	900	900
5	33	1089	31	961	1023
6	24	576	22	484	528
7	18	324	20	400	360
8	25	625	25	625	625
9	22	484	24	576	528
10	17	289	17	289	289
$\Sigma$	264	7512	262	7336	7413

En el Cuadro 6, se presenta la simulación de dos aplicaciones de un instrumento a una muestra de 10 sujetos, con intervalo de dos semanas entre la primera y la segunda aplicación. Las Puntuaciones correspondientes a la primera aplicación aparecen debajo de la columna identificada con la letra X; mientras que las calificaciones de la segunda aplicación aparecen debajo de la columna Y. Para los fines de los cálculos necesarios para calcular el coeficiente de confiabilidad se calcularon los valores de X<sup>2</sup>, Y<sup>2</sup> y XY. Luego se aplicó la fórmula para obtener el coeficiente de correlación por el método de los puntajes directos, el cual se expresa en la fórmula siguiente:

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

Donde:

$r$  = es el coeficiente de correlación entre las dos administraciones de la prueba

$N$  = número de sujetos resultado de sumar el producto de cada valor de X por su correspondiente valor en Y

$\sum X$  = suma total de los valores de X (primera aplicación)

$\Sigma Y$  = suma total de los valores de Y (segunda aplicación)

$\Sigma X^2$  = resultado de sumar los valores de X elevados al cuadrado

$\Sigma Y^2$  = resultado de sumar los valores de Y elevados al cuadrado

$(\Sigma X)^2$  = suma total de los valores de X, elevada al cuadrado

$(\Sigma Y)^2$  = suma total de los valores de Y, elevada al cuadrado

Sustituyendo los valores correspondientes en la fórmula, tenemos:

$$r = \frac{10(7413) - (264)(262)}{\sqrt{[10(7512) - (264)^2][10(7336) - (262)^2]}} = \frac{4962}{5057,62} = 0,98$$

Como se puede observar la  $r = 0,98$ . Este resultado indica que existe una correlación “muy alta” entre las puntuaciones de la primera y la segunda medición, lo cual equivale a decir que el instrumento analizado es altamente confiable, en cuanto a la estabilidad de las puntuaciones a través del tiempo.

**La confiabilidad de versiones equivalentes.** Estase utiliza en el caso dos o más pruebas que miden el mismo constructo. Para ello, se preparan dos versiones de la misma prueba y se administran a la misma muestra de sujetos; luego estas dos distribuciones de puntajes son correlacionadas, para estimar el coeficiente de confiabilidad. Dicho coeficiente combina dos tipos de confiabilidad: estabilidad temporal y consistencia de las repuestas a las diferentes muestras de ítemes (ver Helmstadter, 1964).

En relación con la confiabilidad de versiones equivalentes es crucial considerar el concepto de muestreo de contenido; en el sentido de que es necesario estar seguro que las dos pruebas son realmente equivalentes; es decir, están integradas por muestras de ítemes que son representativas del dominio de contenido bajo medición. Al respecto, es conveniente también que ambas pruebas sean presentadas de la misma manera, cubrir la misma cantidad de contenido, tener el mismo número de ítemes y ser equivalentes en

cuanto al nivel de dificultad de los ítemes. Así mismo, se espera que las pruebas sean similares en cuanto a las instrucciones, límite de tiempo, ejemplos ilustrativos y formato de presentación.

El procedimiento para calcular este coeficiente de confiabilidad es similar al utilizado en el caso de la confiabilidad por reaplicación. La única diferencia consiste en que, en el primer caso, usamos la misma prueba en dos ocasiones diferentes para generar las puntuaciones; mientras que en el segundo, usamos dos formas de la misma prueba para obtener las calificaciones.

***La confiabilidad de consistencia interna.*** Este tipo de confiabilidad permite determinar el grado en que los ítemes de una prueba están correlacionados entre sí. Si los diferentes reactivos de un instrumento tienen una correlación positiva y, como mínimo, moderada, dicho instrumento será homogéneo. En consecuencia, se puede definir la homogeneidad como la consistencia en la ejecución en todos los reactivos de la prueba (Loevinger, 1947; Anastasi, 1961). De allí que en una prueba con un alto grado de consistencia interna, el saber como se desempeña una persona en un ítem, nos permite predecir como lo hará en los demás.

Es importante hacer notar que en el caso de la confiabilidad de consistencia interna, el énfasis se pone en las puntuaciones de los sujetos y no en el contenido o el formato de los reactivos. Por lo tanto, si los ítemes del instrumento correlacionan positivamente entre sí, éste será homogéneo, independientemente del tipo de contenido que se haya utilizado. Por el contrario, la prueba será heterogénea si los reactivos no tienen una correlación positiva entre sí, aun cuando aparentemente estén midiendo el mismo rasgo. Como se puede comprender, la distinción entre lo homogéneo y lo heterogéneo no es una dicotomía, sino un continuo.

Por otra parte, la homogeneidad está relacionada con la característica de **unidimensionalidad** de una prueba, la cual indica que el instrumento mide una sola variable (un rasgo) en lugar de una combinación de ellas. Si una prueba es homogénea, podemos suponer que todos los ítemes miden una característica común.

Sin embargo, es concebible que cada reactivo pueda medir una misma combinación de rasgos y, de esa manera, pudieran estar recíprocamente intercorrelacionados. A la inversa, si una prueba no es homogénea, los reactivos o algunos subgrupos de ellos medirán características diferentes. Por consiguiente, la homogeneidad es una característica necesaria, pero no suficiente, de una prueba destinada a medir un rasgo unitario (Brown, 1980).

En las pruebas multidimensionales; es decir, aquellas que miden diferentes aspectos o dimensiones de un constructo, están formadas por subpruebas; por ejemplo, la Prueba de Aptitud Académica (CNU-OPSU) está conformada por dos subpruebas que miden Razonamiento verbal y Habilidad numérica. Este tipo de instrumentos tiende a presentar coeficientes de confiabilidad de consistencia interna de magnitud moderada, indicando que no son totalmente homogéneas, sin embargo, sus respectivas subpruebas si lo son.

Existen diferentes procedimientos para estimar la confiabilidad de consistencia interna. Algunos de los más conocidos son los siguientes: (a) Kuder-Richardson; (b) Alpha de Cronbach; (c) Dos mitades, corregido por la fórmula de Spearman-Brown; y (d) método de Hoyt. Kuder y Richardson (1937) desarrollaron varios modelos para estimar la confiabilidad de consistencia interna de una prueba, siendo uno de los más conocidos la denominada fórmula 20, el cual se representa de la siguiente manera:

$$r_{tt} = \frac{n}{n-1} * \frac{V_t - \sum pq}{V_t}$$

Donde:

$r_{tt}$  = coeficiente de confiabilidad

$n$  = número de ítemes que contiene el instrumento

$V_t$  = varianza total de la prueba

$\sum pq$  = sumatoria de la varianza individual de los ítemes

Para calcular la confiabilidad por el método K-R<sub>20</sub>, se procede así: en primer lugar, para cada ítem se computa **p**, que es la proporción de sujetos que pasaron un ítem sobre el total de sujetos; luego, se computa **q**, que es igual a 1 – p; se multiplica **pq**; y finalmente se suman todos los valores de pq. El resultado obtenido es la sumatoria de la varianza individual de los ítemes, o sea,  $\sum pq$ ; en segundo lugar, se calcula la varianza total de la distribución de calificaciones ( $V_t$ ); y, en tercer lugar, se aplica la fórmula correspondiente.

El modelo de Kuder-Richardson es aplicable en las pruebas de ítemes dicotómicos en los cuales existen respuestas correctas e incorrectas. A continuación se presenta un ejemplo numérico, con el propósito de ilustrar cómo se calcula el coeficiente de confiabilidad por el método K-R<sub>20</sub>, en una prueba de rendimiento académico.

Supongamos que hemos administrado una prueba de cinco ítemes en una muestra de nueve estudiantes, cuyos resultados se presentan en el Cuadro 7.

*Cuadro 7. Matriz de Ítem por Sujeto de los Resultados de una Prueba de Rendimiento Académico*

Sujetos	Ítemes					Total
	1	2	3	4	5	
1	✓	✓	✓	✓	✓	5
2	✓	✓	✓	✓	✓	5
3	x	✓	✓	✓	✓	4
4	✓	x	✓	✓	✓	4
5	x	✓	✓	✓	✓	4
6	✓	✓	✓	✓	x	4
7	✓	✓	✓	x	x	3
8	x	x	✓	✓	✓	3
9	x	✓	✓	x	x	2
p	0,55	0,77	1,0	0,77	0,66	$V_1 = .94$
q (1-p)	0,45	0,23	0,0	0,23	0,34	
pq	0,25	0,18	0,0	0,18	0,22	$\sum pq = .83$

Sustituyendo los valores correspondientes en la fórmula, tenemos:

$$r_{tt} = (K - R_{20}) = \frac{5}{5-1} * \frac{.94-.83}{.94} = 1,25 * .117 = 0,146$$

De acuerdo con el resultado anterior, se concluye que el instrumento en estudio tiene una confiabilidad de consistencia interna “muy baja”. Al respecto, es importante señalar que este tipo de resultados es esperable en los casos cuando se utilizan pruebas cortas ( $n < 10$  ítemes), ya que, como se ha indicado anteriormente, la confiabilidad de una medida es una función directa de su extensión (número de ítemes) de la prueba.

En los casos de la medición de constructos a través de escalas, en los que no existen respuestas correctas ni incorrectas, sino que cada sujeto marca el valor de la escala que mejor representa su respuesta, Cronbach (1951) derivó, a partir del modelo de Kuder-Richardson (1937), una variante que permite estimar la confiabilidad de consistencia interna en estos casos.

La fórmula es como sigue:

$$r_{tt} = \frac{n}{n-1} * \frac{S_t^2 - \sum S_i^2}{S_t^2}$$

Donde:

$r_{tt}$  = coeficiente de confiabilidad

$n$  = número de ítemes

$s_t^2$  = varianza total de la prueba

$\sum s_i^2$  = es la suma de las varianzas individuales de los ítemes

Observe que las dos fórmulas son básicamente iguales. Lo único que varía es la forma de representar la varianza individual de los ítemes en cada caso. A continuación se presenta un ejemplo numérico para ilustrar el procedimiento de cálculo del coeficiente de confiabilidad obtenido por el método Alpha de Cronbach (1951). Supongamos que hemos administrado una escala de actitud hacia la matemática, de 6 ítemes, a una muestra de  $n = 8$  estudiantes de noveno grado. La escala constaba de cinco posiciones en las cuales el valor 5 indicaba la actitud máxima favorable; mientras que el valor 1 denotaba la actitud máxima desfavorable. Los resultados se presentan en el Cuadro 8.

*Cuadro 8. Matriz de ítem por sujeto sobre los resultados de una escala de actitud hacia la matemática en estudiantes de noveno grado*

Sujetos	Ítemes						Total	Total Par	Total Impar
	1	2	3	4	5	6			
A	5	4	3	2	1	1	16	7	9
B	4	1	3	3	1	2	14	6	8
C	5	5	5	5	5	5	30	15	15
D	4	4	2	3	1	1	15	8	7
E	5	2	2	4	3	1	17	7	10
F	5	3	3	2	4	2	19	7	12
G	4	5	2	3	2	3	19	11	8
H	5	5	2	5	2	5	24	15	9
<b>Total</b>	37	29	22	27	19	20	154		
$V_i$	0,27	2,26	1,07	1,41	2,26	2,85			

$$\sum V_i = 10,12$$

$$V_{par} = 13,71$$

$$V_{impar} = 6,78$$

$$s_i^2 = 28,5$$

Sustituyendo los valores correspondientes en la fórmula, tenemos:

$$r_{tt} = \frac{6}{6-1} * \frac{28,5-10,12}{28,5} = 1,2 * 0,6449 = 0,77$$

De acuerdo con el resultado anterior, se concluye que la escala de actitud hacia la matemática tiene una confiabilidad de consistencia interna alta, aunque no lo suficientemente alta como para ser considerada aceptable en el caso de las escalas de actitud, cuyo coeficiente de homogeneidad siempre se espera que caiga en la categoría de “muy alto” (entre 0,81 y 1.0).

La estimación de la confiabilidad a través del método de **dos mitades**, consiste en: (a) dividir los ítemes de la prueba en dos parte iguales; (b) correlacionar las puntuaciones totales de las dos mitades; (c) multiplicar el

coeficiente obtenido por 2 y dividir por el término 1 (uno) más la correlación de las dos mitades, como se expresa en la fórmula que aparece más abajo.

$$r_{tt} = \frac{2r_{hh}}{1+r_{hh}}$$

Donde:

$r_{tt}$  = coeficiente de confiabilidad

$r_{hh}$  = correlación entre las dos mitades

A continuación se presenta un ejemplo numérico con el propósito de ilustrar el procedimiento de cálculo del coeficiente de confiabilidad de consistencia interna por el método de dos mitades, corregido por la fórmula de Spearman-Brown. En el ejemplo anterior, presentado en el Cuadro 8, la correlación de las dos mitades de la prueba fue de 0,72. Sustituyendo los valores correspondientes en la fórmula tenemos:

$$r_{tt} = \frac{2(0,72)}{1+0,72} = \frac{1,44}{1,72} = 0,77$$

El coeficiente de confiabilidad obtenido es de  $r_{tt} = 0,77$ , el cual puede ser considerado como “alto”. Si se compara este valor con el obtenido por el método Alpha de Cronbach ( $r_{tt} = 0,72$ ), podemos observar que el método de dos mitades tiende a inflar ligeramente el coeficiente de confiabilidad. Un modelo alternativo para calcular la confiabilidad por el método de las dos mitades fue desarrollado por Rulón (1939). En este caso sólo se requiere: (a) calcular la diferencia de varianza de las dos mitades para cada sujeto; (b) dividir por la varianza de los puntajes totales; y (c) restar esta proporción de la unidad (1,00). La fórmula es:

$$r_{tt} = 1 - \frac{S_d^2}{S_T^2}$$

Donde:

$r_{tt}$  = es el coeficiente de confiabilidad

$s_d^2$  = es la varianza diferencial entre las dos mitades del test

$s_t^2$  = es la varianza total de la prueba

De acuerdo con los datos reportados en el ejemplo presentado en el Cuadro 7, se tiene que:

$$S_d^2 = V_{par} - V_{impar}$$

Sustituyendo los valores correspondientes tenemos:

$$s_d^2 = 13,71 - 6,78 = 6,93$$

$$s_t^2 = 28,5$$

Aplicando la fórmula de Rulón tendríamos:

$$r_{tt} = 1 - \frac{6,93}{28,5} = 1 - 0,24 = 0,76$$

El coeficiente de confiabilidad de consistencia interna fue de  $r_{tt} = 0,76$ , el cual puede ser considerado como alto. Observe que el valor obtenido a través del método de Rulón se aproxima bastante a la magnitud del coeficiente obtenido mediante el método Alpha de Cronbach.

La confiabilidad de consistencia interna también puede ser estimada mediante el método de Hoyt (1941). Este consiste en aplicar el método del Análisis de Varianza para estimar la confiabilidad de consistencia interna. Para tal fin, se procede así: (a) se elabora una matriz de ítem por sujeto; (b) se calcula la varianza total de la distribución de puntajes totales; (c) se computan los diferentes componentes de la varianza total: varianza debida a los sujetos ( $V_s$ ), varianza debida a los ítemes ( $V_i$ ) y varianza residual o de error ( $V_e$ ); (d) se divide  $V_s/V_e$  para obtener la  $F_s$ ; y  $V_i/V_e$  para obtener la  $F_i$ ; (e) se

determina la significación estadística de cada F. La fórmula de Hoyt es la siguiente:

$$r_{tt} = 1 - \frac{S_e^2}{S_s^2}$$

Donde:

$s_e^2$  = varianza de error

$s_s^2$  = varianza debida a los sujetos

A continuación utilizaremos un ejemplo numérico para ilustrar cómo se calcula el coeficiente de confiabilidad por el método de Hoyt. Para tal fin, utilizaremos los datos del Cuadro 8. Los datos del resumen del análisis de varianza se presentan en el Cuadro 9.

*Cuadro 9. Resumen de ANAVA*

vFuente de variación	SC	gl	MC	F	p
Ítemes	28,92	5	5,78	5,35	< 0,01
Sujetos	33,25	7	4,75	4,40	< 0,01
Residuo	37,75	35	1,08		
Total	99,92	47			

Aplicando la fórmula de Hoyt tenemos:

$$r_{tt} = 1 - \frac{1,08}{4,75} = 0,77$$

De acuerdo con el resultado anterior, se concluye que el instrumento en estudio tiene un coeficiente de confiabilidad alto. En resumen, se podría decir que la escala de actitud hacia la matemática tiene una confiabilidad de consistencia interna promedio de  $r_{tt} = 0,76$  (ver Cuadro 10).

*Cuadro 10. Coeficientes promedio y parciales de confiabilidad de consistencia interna estimada por varios métodos*

Método	$r_{tt}$
Alpha de Cronbach	0,72
Spearman-Brown	0,77
Rulón	0,76
Hoyt	0,77
Promedio	0,76

Es importante señalar que los diferentes métodos empleados para estimar la confiabilidad de consistencia interna dan resultados parecidos. En consecuencia, cuando se desarrolla una prueba no es necesario utilizar todos estos métodos, sino que es suficiente con utilizar sólo uno de ellos.

### **Interpretación del Coeficiente de Confiabilidad**

Como ya se ha mencionado anteriormente, la confiabilidad de un instrumento se expresa mediante un coeficiente de correlación:  $r_{tt}$ , que teóricamente significa correlación del test consigo mismo. Sus valores oscilan entre cero (0) y uno (1). Una manera práctica de interpretar la magnitud de un coeficiente de confiabilidad puede ser guiada por la escala que se presenta en el Cuadro 11.

*Cuadro 11. Interpretación de la Confiabilidad*

Rangos	Magnitud
0,81 a 1,00	Muy Alta
0,61 a 0,80	Alta
0,41 a 0,60	Modearada
0,21 a 0,40	Baja
0,01 a 0,20	Muy Baja

Por lo general, un coeficiente de confiabilidad se considera aceptable cuando está por lo menos en el límite superior (0,80) de la categoría “Alta”. No obstante, no existe una regla fija para todos los casos. Todo va a depender del tipo de instrumento bajo estudio, de su propósito y del tipo de confiabilidad de que se trate. Por ejemplo, un coeficiente de confiabilidad de consistencia interna para una escala de actitud, nunca debería estar por debajo del límite inferior de la categoría “muy alto”, o sea, el valor de  $r_{tt} = 0,81$  para ser considerado como aceptable. Sin embargo, en el caso de una prueba de rendimiento académico, la literatura reporta coeficientes que varían entre 0,61 y 0,80 (ver Thorndike, 1989; Magnusson, 1983). En el caso de instrumentos con coeficientes de confiabilidad moderados, una manera de saber hasta dónde los mismos pueden ser aceptables, consiste en comparar la desviación estándar de la distribución de puntajes ( $S_y$ ) con el error estándar de medición (EEM), cuya fórmula se indica más abajo.

$$E_{EM} = S_y \sqrt{1 - r_{tt}}$$

Donde:

$E_{EM}$  = error estándar de medición

$S_y$  = desviación estándar de la distribución de puntajes totales

$r_{tt}$  = coeficiente de confiabilidad

En estos casos se recurre al criterio del error estándar de medición para

decidir sobre la aceptabilidad de un coeficiente de confiabilidad moderado, se requiere que se cumpla la condición de que  $S_y > E_{EM}$ . Veamos un ejemplo; supongamos que la confiabilidad de un instrumento fue de 0,58 y tenemos duda acerca de su aceptación. Asumamos, además que  $S_y = 1,5$ . Si aplicamos la fórmula para estimar el valor del  $E_{EM}$ , tenemos:

$$E_{EM} = 1,5\sqrt{1 - 0,58} = 0,97$$

Como se puede observar, en el caso anterior el  $E_{EM}$  (0,97) no excede el valor de la  $S_y$  (1,5); es decir, que se cumple la condición de aceptabilidad señalada anteriormente ( $S_y > E_{EM}$ ). En consecuencia, el instrumento a pesar de su nivel de confiabilidad podría ser utilizado, pero con reservas. En tal sentido, el investigador (o evaluador) tendría que ser muy cauteloso en la interpretación de sus resultados.

### **Mejoramiento de la Confiabilidad de una Prueba**

La confiabilidad de un instrumento está en relación directa con su extensión. Esto se explica porque en la medida que la prueba está formada por pocos ítems ( $n < 10$ ) el error de medición aumenta y, en consecuencia, la confiabilidad tiende a bajar. Es decir, cuando la prueba contiene una muestra grande de ítems ( $n > 30$ ) que son representativos del dominio que se pretende medir, aumenta la probabilidad de acercarnos a la varianza verdadera del instrumento y, de esta manera, se incrementa la confiabilidad.

Además, la confiabilidad de una prueba puede ser mejorada, siguiendo el principio del **maxmicon** sugerido por Kerlinger y Lee (2002), el cual consiste en: maximizar la varianza debida a las diferencias individuales y minimizar la variación de error. En resumen, el procedimiento recomendado por Kerlinger es el siguiente:

Primero, escribir los ítems de manera clara, sin ambigüedad. Un ítem

ambiguo puede ser interpretado de muchas maneras, lo cual contribuye a aumentar el error de la varianza y, por lo tanto, disminuye la confiabilidad de la prueba.

Segundo, escribir claramente las instrucciones, contribuye a reducir el error de medición. Las instrucciones ambiguas y o poco claras contribuyen a aumentar el error de la varianza.

Tercero, usar condiciones estándares y bien controladas para administrar la prueba. Los factores internos (referidos al sujeto que toma la prueba) como los externos (asociados a las condiciones bajo las cuales se toma la prueba, contribuyen a aumentar el error de medición.

### **Importancia de la Confiabilidad**

Para que los resultados de un instrumento puedan ser interpretables; es decir, para que tengan significado y valor heurísticos, es necesario que los mismos sean confiables. No es posible determinar la relación entre dos o más variables si los instrumentos utilizados para medirlas son poco confiables. Este tipo de instrumento, a menudo, inducen al investigador a interpretaciones erradas de sus resultados. Por ejemplo, un experimento educativo, en el cual se ha evaluado el efecto de una nueva estrategia instruccional, pudiera parecer poco efectivo, aun habiendo tenido un impacto altamente significativo en el aprendizaje de los alumnos, simplemente porque los instrumentos utilizados para medir las variables dependientes eran poco confiables.

La confiabilidad, aun cuando no es la característica más importante de un instrumento de medición, requiere se le preste toda la atención que sea necesaria. Ciertamente, una alta confiabilidad, por si sola, no garantiza “buenos” resultados científicos. Pero, no puede haber “buenos” resultados científicos sin instrumentos confiables. En síntesis, la confiabilidad es una

condición necesaria, pero no suficiente para obtener resultados de investigación que sean científicamente valiosos y socialmente útiles.

## Validez

En la sección anterior, nuestro interés se centró en determinar hasta dónde los resultados de un instrumento de medición son estables a través del tiempo, cuando éste ha sido utilizado varias veces con los mismos sujetos y bajo las mismas condiciones de aplicación. En esta sección, nos interesa estudiarla exactitud con que pueden hacerse mediciones significativas y adecuadas con un instrumento, en el sentido de que mida realmente el rasgo que pretende medir. Esta propiedad o característica de un instrumento de medición recibe el nombre de **validez**. Es decir, en sentido general, la validez de un instrumento tiene que ver con las preguntas siguientes: ¿qué miden los puntajes del test? y ¿qué predicen dichas puntuaciones?(Guilford, 1954; Nunnally, 1967; Anastasi, 1976; Magnusson, 1982).

Por ejemplo, si un rasgo, como la inteligencia académica, es significativo o relevante para el éxito en el aprendizaje formal, construimos un instrumento para medir dicho rasgo y lo usamos en las tareas de diagnóstico, selección y de orientación escolar o académica. Pero, obviamente, es necesario que el instrumento mida el rasgo que hemos encontrado relevante en la situación donde el mismo va ser utilizado y no alguna otra característica.

La validez de un instrumento, por lo general, no constituye un problema en el caso de la medida de los objetos físicos, tales como longitud, peso, capacidad. Por supuesto que la estatura de una persona se mide con una cinta métrica y el peso de un objeto, con una balanza (siempre que la balanza funcione adecuadamente). Sin embargo, con los métodos usados para medir

variables psicoeducativas, es necesario probar empíricamente que el instrumento es válido en todos los casos.

Cuando elaboramos una escala para medir la actitud de los docentes hacia la innovación educativa, debemos probar que los puntajes de la escala realmente distinguen entre aquellos docentes que tienen una actitud favorable hacia la innovación y aquellos cuya actitud es desfavorable. Algunas veces, puede parecer obvio que un instrumento mida un determinado rasgo; sin embargo, pudiera ocurrir que cuando se le someta a una evaluación empírica, el instrumento esté midiendo algún rasgo distinto. De allí que sea necesario tener presente que la validez no es materia de presunción sino de demostración empírica.

Cuando estimamos la validez de un instrumento, necesitamos saber qué característica deseamos que prediga. Este rasgo se llama **variable criterio**. Nos interesa saber qué tan bien corresponden las posiciones de los individuos en las distribución de los puntajes obtenidos con respecto a sus posiciones en el continuo que representa la variable criterio. Por lo general, la validez es estimada a través de una correlación entre los puntajes de una medida y las puntuaciones de la variable criterio, la cual predice. Este índice se denomina coeficiente de validez.

Un instrumento no tiene un coeficiente fijo de validez que sirva para cualquier propósito y para cualquier grupo de individuos. La validez de un instrumento varía de acuerdo con el propósito con que se use y el grupo dentro del cual discrimina. Por ejemplo, si una prueba de aptitud académica predice el rendimiento (variable criterio) de los estudiantes en el primer semestre de la universidad, es de esperarse que la mayoría de los estudiantes que obtuvieron altas calificaciones en la prueba de aptitud, también sean los que obtengan más éxito en dicho lapso académico. Sin embargo, el hecho de

que la prueba de aptitud académica haga una predicción razonable del rendimiento universitario, no garantiza que también pueda predecir la motivación al logro de dichos estudiantes, o quiénes se graduarán al final de la carrera, ni muchos menos quiénes tendrán éxito en sus respectivos campos profesionales.

### **Tipos de Validez**

La validez como la confiabilidad de un instrumento, a pesar de tener ambas sus respectivas definiciones teóricas genéricas, en la práctica, dado que estas características pueden adoptar diferentes significados, es necesario adjetivarlas cada vez que se haga referencia a ellas. Es decir, no es suficiente con que se diga que un instrumento determinado tiene una alta confiabilidad o validez. Por qué el lector especializado inmediatamente se preguntaría ¿qué tipo de confiabilidad? o ¿qué tipo de validez? Es necesario ser específico al usar estos términos. De esta manera se podría decir, por ejemplo, este instrumento tiene una alta confiabilidad de consistencia interna o esta prueba tiene una alta validez predictiva.

En el caso particular de la validez, cuando investigamos esta característica en un instrumento determinado, intentamos responder tres tipos de cuestiones, que aluden a igual número de tipos de validez. Estas cuestiones son:

1. ¿Cuán representativo es el comportamiento elegido como muestra del universo que se intenta representar?
2. ¿Qué significado tiene el comportamiento con respecto a los atributos del individuo que son de interés para la medición?
3. ¿Hasta dónde se puede predecir el rendimiento del sujeto o su aprendizaje en un programa de entrenamiento (o hasta dónde se puede

anticipar su nivel de desempeño en el trabajo), a partir de su ejecución en la prueba?

Estos tres tipos de preguntas corresponden a lo que en la literatura especializada se ha denominado respectivamente como **validez de contenido**, **validez de constructo** y **validez predictiva**. A continuación se presenta una descripción de cada una de ellas, así como sus respectivos procedimientos para estimar los diferentes índices de validez.

### **Validez de Contenido**

A través de la validez de contenido se trata de determinar hasta dónde los ítemes de un instrumento son representativos del dominio o universo de contenido de la propiedad que se desea medir. Esta definición, aunque bien en teoría, presenta dificultades prácticas, ya que es imposible extraer muestras aleatorias de reactivos de un universo de contenido, puesto que este existe sólo conceptual o teóricamente.

Algunos autores han tratado de resolver el problema, de la validez de contenido, generando grandes cantidades de ítemes de un dominio determinado, para luego obtener muestras representativas de dicho universo, a los fines de integrar un instrumento; sin embargo, en la opinión de autores como Kerlinger y Lee (2002), la validez de contenido de tales conjuntos, independientemente de qué tan grandes y tan “buenos” sean los reactivos, es siempre dudosa y, por tanto, cuestionable. ¿Cómo lograr, entonces, una estimación de la validez de contenido de un instrumento?

A diferencia de otros tipos de validez, la de contenido no puede ser expresada cuantitativamente, a través de un índice o coeficiente; ella es más bien una cuestión de juicio. Es decir, la validez de contenido, por lo general, se estima de manera subjetiva o intersubjetiva. El procedimiento más comunmente empleado para determinar este tipo de validez, es el que se

conoce con el nombre de **juicios de expertos**, para lo cual se procede de la siguiente manera:

1. Se seleccionan dos jueces o más expertos a los fines de juzgar, de manera independiente, la “bondad” de los ítemes del instrumento, en términos de la relevancia o congruencia de los reactivos con el universo de contenido, la claridad en la redacción y la tendenciosidad o sesgo en la formulación de los ítemes.

2. Cada experto recibe suficiente información escrita acerca de: (a) el propósito de la prueba; (b) conceptualización del universo de contenido; (c) plan de operacionalización o tabla de especificaciones (en el caso de las pruebas de rendimiento académico).

3. Cada juez recibe un instrumento de validación en el cual se recoge la información de cada experto. Dicho instrumento normalmente contiene las siguientes categorías de información por cada ítem: congruencia ítem-dominio, claridad, tendenciosidad y observaciones (ver Figura 1)

Ítem	Congruencia		Claridad		Tendenciosidad		Observaciones
	Si	No	Si	No	Si	No	
1	<input type="checkbox"/>	_____					
2	<input type="checkbox"/>	_____					
3	<input type="checkbox"/>	_____					
4	<input type="checkbox"/>	_____					
5	<input type="checkbox"/>	_____					
6	<input type="checkbox"/>	_____					
7	<input type="checkbox"/>	_____					
8	<input type="checkbox"/>	_____					
9	<input type="checkbox"/>	_____					
10	<input type="checkbox"/>	_____					
11	<input type="checkbox"/>	_____					

*Figura 1. Planilla de Validación*

4. Se recogen y analizan los instrumentos de validación y se toman las decisiones siguientes: (a) los ítemes donde hay un 100 por ciento de coincidencia favorable entre los jueces (los ítemes son congruentes, están escritos claramente y no son tendenciosos) quedan incluido en el instrumento; (b) los ítemes donde hay un 100 por ciento de coincidencia desfavorable entre los jueces, quedan excluidos del instrumento; y (c) los ítemes donde sólo hay coincidencia parcial entre los jueces deben ser revisados, reformulados, si es necesario, y nuevamente validados.

## **Validez de Constructo**

La validez de constructo intenta responder la pregunta ¿hasta dónde un instrumento mide realmente un determinado rasgo latente o una característica de las personas y con cuánta eficiencia lo hace? Esta pregunta tiene sentido, particularmente en los instrumentos que se utilizan en la investigación psicoeducativa, ya que, como se ha mencionado anteriormente, en este campo hacemos mediciones indirectas de ciertas variables internas del individuo que denominamos constructos. En consecuencia, es necesario que podamos mostrar evidencia de que, efectivamente, el instrumento mide el rasgo o constructo que pretende medir.

Gronlund (1976) señala que la validez de constructo interesa cuando queremos utilizar el desempeño de los sujetos con el instrumento para inferir la posesión de ciertos rasgos o cualidades psicológicas. Por ejemplo, en lugar de hablar de los puntajes que una persona obtuvo en un instrumento determinado, podemos hablar de: (a) la actitud hacia la matemática; (b) la satisfacción o disfrute con el aprendizaje de la matemática; y (c) la valoración o significado de la matemática. Todas éstas son cualidades hipotéticas llamadas construcciones, cuya existencia se supone para explicar la conducta en muchas y diferentes situaciones específicas.

Para estudiar la validez de constructo de un instrumento es necesario que exista una conceptualización clara del rasgo bajo estudio, con base en una “teoría” determinada. Esta nos permitirá tener una idea clara acerca de cómo se manifiesta el atributo bajo estudio, qué tipo de rendimiento, en la vida diaria, se facilitan por la posesión del atributo, cuáles subgrupos de la población lo poseen en alto o bajo grado y qué condiciones favorecen o impiden su expresión. La teoría del atributo, también sugiere las tareas de prueba que son las apropiadas para hacer aflorar el atributo. Además, la teoría

sugiere las clases de evidencias que deben considerarse para evaluar las bondades de la prueba bajo estudio, si es que de hecho hace aflorar o depende directamente del constructo que se analiza.

Cronbach (1960) ha sugerido los pasos siguientes para establecer la validez de constructo: (a) identificar las construcciones que pudieran explicar la ejecución en el instrumento; (b) formulación de hipótesis comprobables a partir de la teoría que enmarca a cada construcción; y (c) recopilación de datos para probar estas hipótesis. Estas hipótesis se pueden enunciar en relación con cualquiera de los siguientes tipos generales de evidencia (Thorndike, 1989; Helmstadter, 1964):

1. La relación entre la conceptualización teórica del instrumento y su estructura factorial. Por ejemplo, en el desarrollo de la escala AC-2000, que mide el Auto-concepto de los alumnos de sexto grado, Ruiz (1988) comprobó, a través de la técnica del **Análisis Factorial**, que las tres dimensiones hipotetizadas en la escala (Autoconcepto Social, Autoconcepto Personal y Autoconcepto Escolar) tenían soporte empírico en los datos (ver Cuadro 12)

*Cuadro 12. Estructura factorial de la Escala AC-2000*

Item	Factor 1 A. Social	Item	Factor 2 A. Escolar	Item	Factor 3 A. Personal
3	0,51	4	0,52	1	0,43
6	0,63	8	0,43	2	0,48
16	0,49	12	0,45	5	0,58
22	0,48	25	0,54	7	0,43
26	0,64	28	0,57	9	0,47
27	0,54	29	0,55	10	0,56
		30	0,40	13	0,41

2. La información correlacional, esto es, que muestre las correlaciones entre la prueba bajo estudio y otras medidas (pruebas o eventos en la vida real) que reflejen o dependan del atributo en cuestión. Por ejemplo, en el mismo estudio de Ruiz (1988) se encontró una correlación de 0,79 entre el instrumento AC-2000 y la escala de Autoestima de Coopersmith (1959).

También se pueden formular y evaluar hipótesis que planteen relaciones con otras variables en las que teóricamente: (a) es esperable una correlación positiva con el constructo bajo estudio, conocido en la literatura con el nombre de **validez convergente**; y (b) no es esperable ningún tipo de relación con el constructo objeto de validación, conocida como **validez discriminante** (Campbell y Fiske, 1959).

3. Los datos sobre las diferencias entre grupos, comparando las puntuaciones de prueba de los subgrupos que se puede esperar difieren en el nivel del atributo. Por ejemplo, en el desarrollo de una escala de Autoconcepto, se podría probar la hipótesis de que los estudiantes con alto y bajo auto-concepto académico difieren en rendimiento escolar.

4. La información que muestre los efectos de tratamiento o intervenciones experimentales que se puede esperar influyan en la expresión del atributo. Por ejemplo, existen evidencias en la literatura que brindan soporte a la hipótesis de que una estrategia instruccional centrada en la retroalimentación positiva y crítica en las pruebas de evaluación formativa, de estudiantes de sexto grado, mejora el Auto-concepto de dichos alumnos (ver Sánchez de Hurtado, 1994).

5. Consistencia interna. En este sentido, podríamos predecir correlaciones altas entre ítems debido a que todos ellos supuestamente miden el mismo constructo: Auto-concepto.

### **Validez Predictiva**

La palabra predicción o predictivo normalmente se la asocia con visión o anticipación de futuro. En este sentido, cuando estudiamos la validez predictiva de un instrumento lo que nos interesa es determinar hasta dónde podemos anticipar el desempeño futuro de una persona en una actividad determinada, a partir de su ejecución actual en dicho instrumento; por ejemplo, se podría estudiar, hasta dónde la Prueba de Aptitud del Consejo Nacional de Universidades predice el éxito académico de los estudiantes en los primeros semestres universitarios, o anticipar el desempeño futuro de un vendedor a partir de su ejecución en un test de inteligencia social.

En consecuencia, la validez predictiva, también llamada validez de criterio externo o validez empírica, se estudia comparando los puntajes de un instrumento (variable independiente) con una o más variables externas (variables dependientes) denominadas variables criterio. Se asume que tales criterios, indicadores del desempeño futuro, están teórica y lógicamente relacionados con el rasgo representado en el instrumento bajo estudio. Esta comparación entre los puntajes de la variable en estudio y los de la variable

criterio se expresa a través de un coeficiente de correlación, el cual se interpreta como un índice de validez. Entre más alta sea la correlación entre una medida o medidas de aptitud académica y el promedio de notas, tomado como variable criterio, mejor será la validez predictiva de la prueba de aptitud académica.

Es importante destacar que la validez predictiva suele estar asociada con problemas y resultados prácticos; es decir, el interés no es tanto en lo que está detrás del desempeño en la prueba, sino más bien en ayudar a resolver problemas prácticos y tomar decisiones. Muchos de estos problemas y toma de decisiones están relacionados con la evaluación, selección y asignación de personas para diferentes actividades (estudio, trabajo, deporte, arte, etc).

No obstante, determinar la validez predictiva de un instrumento tiene sus problemas, los cuales en buena medida están relacionados con la calidad del criterio externo seleccionado, ya que si éste no es significativo o relevante con respecto a lo que mide el instrumento bajo estudio, la validez predictiva no podrá quedar evidenciada, pero no porque efectivamente la prueba no tenga un determinado tipo de validez predictiva, sino porque el criterio utilizado, para probar dicha propiedad del instrumento, no fue el apropiado.

Es decir, la obtención de criterios apropiados para validar una prueba a veces tiene sus dificultades; por ejemplo, ¿cuál de los siguientes podría ser el mejor criterio para predecir el nivel de eficiencia de un docente? ¿el nivel de dominio que los alumnos tienen de los objetivos instruccionales? ¿el promedio de calificaciones? ¿la tasa de estudiantes aprobados o promovidos? ¿la claridad con que el docente expone sus clases?, ¿el dominio que el docente tiene del contenido de la asignatura que enseña? ¿la responsabilidad del docente en el cumplimiento de las tareas inherentes a su cargo?

Como se pudo observar en el párrafo anterior, todos los criterios

mencionados pudieran estar, de alguna manera, relacionados con la eficiencia de un docente; el problema está en cómo determinar el más relevante, o los más relevantes, porque puede darse el caso de que el desempeño futuro que se aspira predecir, a partir de la ejecución actual en una prueba, se exprese no a través de un criterio sino de varios, como pareciera ser el caso, ya mencionado, de la eficiencia de un docente.

Lo anterior indica que, en algunos casos, existen criterios múltiples, como también pueden existir los predictores múltiples. De hecho, dada la complejidad multifactorial que determina el comportamiento humano, en cualquier actividad donde se desempeñe, la tendencia actual, en los procedimientos de selección (de estudiantes, trabajadores, deportistas, artistas, etc) es a utilizar predictores múltiples (v.g. baterías de pruebas), más que instrumentos únicos. De la misma manera, y por la misma razón, cada vez más se utilizan criterios múltiples como indicadores del desempeño, más que criterios simples o únicos.

### **Factores que Afectan la Validez de una Prueba**

La validez de los resultados de un instrumento de medición puede verse afectada por diferentes factores, los cuales pueden estar asociados tanto al proceso de elaboración de la prueba misma, como a otros factores relacionados con los procedimientos de administración, calificación e interpretación de los resultados. A continuación se presentan algunos de estos factores con la intención deliberada de que, en la medida de lo posible, sean evitados en el proceso de construcción de los instrumentos. Entre estos factores están:

1. *Construcción del instrumento.* Estos factores pueden influir en que las preguntas del instrumento no funcionen en la forma prevista y, por lo tanto, contribuyen a minar la validez de los resultados; algunos de

ellos son: (a) instrucciones imprecisas o vagas; (b) estructura sintáctica de la oración demasiado difícil; (c) preguntas con niveles de dificultad inapropiados; (d) preguntas que sugieren la respuesta; (e) ambigüedad en la formulación de los reactivos, lo cual lleva a diferentes interpretaciones de los mismos; (f) pruebas demasiado cortas; (g) ítemes incongruentes con el dominio o universo de contenido; (h) ordenamiento inadecuado de los ítemes; (i) patrón identificable de respuestas, particularmente en las preguntas de selección.

2. *Administración y calificación de la prueba.* Entre estos factores se encuentran: tiempo insuficiente para responder, ayuda adicional a algunos sujetos, más allá de las instrucciones generales y de los ejercicios de práctica, uso de “chuletas” o ayudas escritas, por parte de los sujetos, uso de la subjetividad de la puntuación de las preguntas.

3. *Respuestas de los sujetos.* Algunos de estos factores son: bloqueo de los sujetos para responder, debido a situaciones emocionales y las respuestas formuladas al azar.

4. *Naturaleza del grupo y del criterio.* Como ya se ha mencionado, la validez es siempre específica con respecto a un grupo en particular. Por lo tanto, si una prueba de comprensión de lectura, por ejemplo, que ha sido diseñada y validada para alumnos de sexto grado, pudiera carecer de validez de contenido y predictiva, para medir el nivel de comprensión de lectura en estudiantes de noveno grado.

Por otra parte, es sumamente importante prestar atención al criterio que va a ser seleccionado como validador del instrumento, ya que, como se mencionó anteriormente, el uso de un criterio inapropiado puede disfrazar o distorsionar la validez de la prueba.

### **La Validez como Varianza Común**

En la sección anterior se señaló que la confiabilidad, genéricamente hablando, podía ser interpretada como la correlación del instrumento consigo mismo, y que, además, podía ser definida como la proporción de la varianza verdadera ( $V_v$ ) sobre la varianza total; o sea como:

$$r_{tt} = \frac{V_v}{V_t}$$

De la misma manera, la validez puede ser interpretada como la relación que un instrumento, bajo estudio, tiene con otro que se utiliza como criterio y puede ser definida como el porcentaje de varianza común que dicho instrumento comparte con el criterio, como se expresa en la fórmula siguiente:

$$Val = \frac{V_{co}}{V_t}$$

Donde:

$V_{al}$  = validez del instrumento

$V_{co}$  = varianza común que el instrumento comparte con el criterio

$V_t$  = varianza total del instrumento bajo estudio

Pero, ¿Qué se entiende por varianza común? ¿Cómo se relaciona la varianza verdadera ( $V_v$ ) con la varianza común ( $V_{co}$ )? ¿Qué se entiende por varianza específica de una prueba? ¿Qué es la varianza de error? Para intentar responder estas interrogantes presentaremos algunas nociones básicas e introductorias de la teoría factorial.

De acuerdo con esta teoría, cuando dos pruebas están relacionadas, como ocurre en el caso de la validación de un instrumento, pueden ser identificados varios tipos de varianzas, como son: (a) la *varianza de factor común*; es decir, el porcentaje de varianza que un instrumento comparte con otro; (b) la *varianza específica*, que es la varianza sistemática de una medida

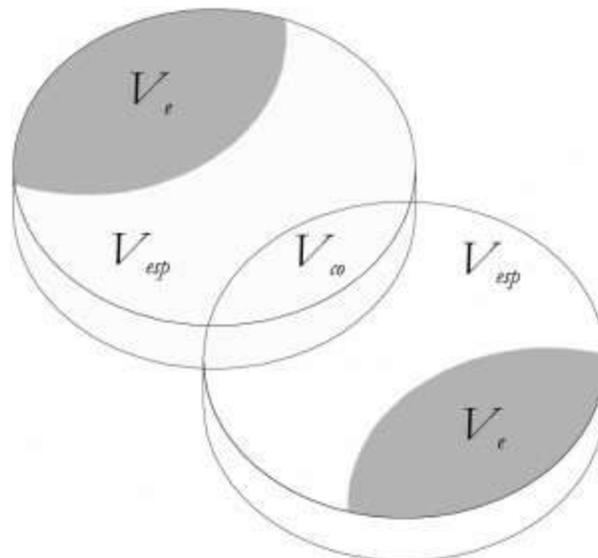
que no es compartida con otra; y (c) la *varianza de error*, que es una varianza aleatoria (no sistemática) debido a factores inherentes a la construcción de la prueba, a sus condiciones de aplicación o a situaciones personales presentadas por el sujeto en el momento de responder el instrumento.

Es decir, si una prueba mide habilidades que también miden otras pruebas, se tiene una varianza de factor común; por el contrario, si la prueba mide una habilidad que no mide ninguna otra prueba, se tiene una varianza específica (Kerlinger y Lee 2002).

De acuerdo con esto, la varianza total de una medida está integrada por varios componentes, a saber: varianza de factor común, varianza específica y varianza de error, lo cual puede ser expresado mediante la siguiente ecuación:

$$V_t = V_{co} + V_{es} + V_e$$

Esta relación puede ser representada también gráficamente, como se indica a través de la Figura 2.



*Figura 2. Componentes de la varianza de dos instrumentos (Adaptado de Kerlinger y Lee, 2002)*

En cuanto a la relación entre la varianza verdadera ( $V_v$ ) y la varianza común, observemos que la varianza total, por una parte, puede ser expresada como la sumatoria de la varianza verdadera, más la varianza de error, como se presenta en la siguiente ecuación:

$$V_t = V_v + V_e$$

Pero, al mismo tiempo, la varianza total puede ser expresada como la sumatoria de la varianza de factor común, más la varianza específica, más la varianza de error, como se señala en la siguiente ecuación:

$$V_t = V_{co} + V_{es} + V_e$$

( $V_v$ ) con la varianza común ( $V_{co}$ )

De lo anterior se deduce que la varianza verdadera ( $V_v$ ) está integrada por la varianza común ( $V_{co}$ ), más la varianza específica ( $V_{es}$ ), como se presenta en la Figura 3.

$$\text{Varianza Total} = \underbrace{\text{Varianza Común} + \text{Varianza Específica}}_{\text{Varianza Verdadera}} + \text{Varianza de Error}$$

*Figura 3. Componentes de la Varianza*

### **Relación entre Validez y Confiabilidad**

Si la confiabilidad puede ser interpretada como la proporción de la varianza

verdadera ( $V_v$ ) con respecto a la varianza total del instrumento, como se expresa en la ecuación:

$$r_{tt} = \frac{V_v}{V_t}$$

Y, por otra parte, la validez es definida como la proporción de la varianza común con respecto a la varianza total, como se indica en la ecuación:

$$Val = \frac{V_{co}}{V_t}$$

Si sabemos, además, que la varianza total ( $V_t$ ) puede ser descompuesta en sus componentes, como se expresa en la fórmula:

$$\frac{V_t}{V_t} = \frac{V_{co}}{V_t} + \frac{V_e}{V_t}$$

Pero, a su vez, la varianza verdadera ( $V_v$ ) está compuesta por la varianza común, más la varianza específica ( $V_{es}$ ), la ecuación anterior puede ser reformulada como sigue:

$$\frac{V_t}{V_t} = \frac{V_{co}}{V_t} + \frac{V_{es}}{V_t} + \frac{V_e}{V_t}$$

De lo anterior se deduce que la proporción de varianza que comprende la interpretación de la confiabilidad  $r_{tt} = (V_{co} + V_{es})/V_t$ , es mayor que la proporción de varianza incluida en la interpretación de la validez  $Val = V_{co}/V_{es}$ . No obstante, desde el punto de vista teórico, pudiera darse la circunstancia en que un instrumento sólo tuviera  $V_{co} + V_e$  y no tuviera  $V_{es}$ , en este caso la proporción de varianza que define la confiabilidad sería igual a la que define a la validez. Por lo tanto, en términos de magnitudes es de

esperarse que, bajo condiciones normales, el coeficiente de confiabilidad de un instrumento siempre sea mayor o igual que su coeficiente de validez.

Ilustremos las relaciones anteriores con un ejemplo numérico. Supongamos que hemos diseñado una prueba para medir Vocabulario (PVO) en sexto grado y que hemos seleccionado una prueba de Comprensión de Lectura (PCL) como criterio externo de validación y encontramos que la varianza común entre PVO y PCL es igual a 0,35; mientras que la que comparte PCL con PVO es de 0,30.

Supongamos, además, que la varianza específica de PVO es de 0,20 y que la varianza de error es de 0,15. En tal situación se cumplen que: (a) las dos pruebas tienen una varianza compartida (Factor común A + B) de 0,65 (coeficiente de validez de PVO) ; (b) la varianza verdadera ( $V_{co} + V_{es}$ ) es de 0,85 (coeficiente de confiabilidad de PVO). Se cumple la tendencia de que en la mayoría de los casos la proporción de varianza que comprende la confiabilidad del instrumento es mayor que la que está implícita en la validez (Ver Cuadro 13 ).

*Cuadro 13. Relación entre los conceptos de confiabilidad y validez de acuerdo con los supuestos de la teoría factorial*

Componentes de la Varianza	Vocabulario (A)	Comprensión de Lectura (B)
Factor Común	.35	.30
Varianza Compartida ( $V_{co}$ )	(A+B)	.65
Varianza Específica ( $V_{es}$ )	.20	
Varianza Verdadera	( $V_{co}+V_{es}$ )	.85
Varianza de Error ( $V_e$ )	.15	

## **Importancia de la Validez**

Una de las principales dificultades que se presenta con la validez de los instrumentos en Educación y Psicología, como ya se ha mencionado antes, es que la medición siempre es indirecta; es decir, debemos inferirla a partir del desempeño que los sujetos tienen en las pruebas. Este hecho en sí mismo representa un problema, ya que ¿cómo sabemos que estamos haciendo la inferencia correcta a partir de los resultados?. Pero, el problema de fondo sigue siendo el de poder asegurarse de que el instrumento mida lo que pretende medir, es decir, que sea válido a los fines propuestos. Muchos de los cuestionamientos que se hacen, hoy día, en la investigación sociológica y psicoeducativa tiene que ver, primero que todo, con la validez de la medida. Obviamente, un instrumento que no tiene una validez demostrada no vale la pena que sea utilizado.

Si comparamos la confiabilidad con la validez, nos damos cuenta que la obtención de la primera puede ser reducida básicamente a una cuestión técnica. Sin embargo, la validez es mucho más que eso. Tiene que ver con el aspecto sustantivo de la ciencia misma. También se relaciona con la epistemología, en tanto que teoría del conocimiento, y con los paradigmas científicos. Particularmente, en lo que se refiere a la validez de constructo, puesto que tiene que ver con la naturaleza de la “realidad” a las propiedades del objeto que se están midiendo (Kerlinger y Lee, 2002).

No obstante, las dificultades prácticas que se presentan para lograr obtener medidas válidas y confiables, dentro del paradigma de la ciencia clásica, en los últimos años se han desarrollado una serie de métodos, técnicas y procedimientos, que facilitan, cada vez más, esta tarea. Pero, más que el manejo de todo este instrumental tecnológico (métodos estadísticos, procedimientos electrónicos, paquetes computarizados, etc), lo más

importante es que el investigador se haga consciente de la necesidad de utilizar instrumentos apropiados, técnicamente bien calibrados, a fin de garantizar la utilidad y significado de los resultados obtenidos.

Queda claro entonces que la construcción de instrumentos de medición no se reduce a la simple presentación de un listado de preguntas en un formato determinado. Construir “buenos” instrumentos de medición es, primero que todo, una tarea técnica, que requiere, por parte del investigador, un entrenamiento específico para acometerla con éxito. Construir un instrumento técnicamente bien hecho implica, en sí mismo, una investigación. De allí que cuando se requiera hacer un estudio (trabajo o tesis de grado, trabajo de ascenso, investigación libre) antes de tomar la seria decisión de construir un instrumento de medición, sin ser un especialista en el área, se debería averiguar previamente acerca de la existencia de la disponibilidad comercial de dicho instrumento en el mercado, o a través de otros investigadores. Si después de esta indagación se llega a determinar que el instrumento no existe y que es indispensable trabajar en el desarrollo del mismo, lo más recomendable sería buscar el asesoramiento técnico especializado correspondiente.

# Construcción de Escalas de Actitudes

La actitud ha sido posiblemente uno de los constructos más estudiados a través de la historia de la psicología. Para algunos autores es el concepto más importante de la psicología social (McGuire, 1969; Berkowitz, 1972). Ha sido tal su importancia que ya en 1918 Thomas y Znaniecki pensaban que la psicología social era el estudio científico de las actitudes, aun cuando hoy día se sabe que este campo de la psicología ha sido objeto de un gran crecimiento y actualmente abarca una gran cantidad y variedad de conductas sociales tanto humanas como animales.

El auge del estudio de las actitudes ha sido paralelo al desarrollo de la sociedad de consumo de Occidente y a los cambios que han ocurrido a nivel mundial durante los últimos 50 años. La diversificación de la producción en masa y la lucha por la competencia en un mercado finito de consumidores ha estimulado el desarrollo de la publicidad, a través de diferentes medios tales como la internet, la televisión, la radio, la prensa y otros, para destacar las cualidades de un producto, bien o servicio, frente a otro y de esta manera inducir el consumo sobre la base de propiciar cambios de actitudes y conductas en la población.

El estudio de las actitudes ha sido también de gran interés en el campo político. Desde hace mucho tiempo las élites gobernantes de los países consideran importante los efectos de la propaganda política en la población, como estrategia para mantener o aumentar la popularidad gubernamental o como vía para alimentar el clientelismo y afianzar una determinada ideología.

Por ejemplo, durante la Segunda Guerra Mundial fue importante la propaganda que las partes involucradas hacían cada una a su favor (Whittaker, 1979).

En efecto para el año de 1930, los investigadores que trabajaban para el ejército norteamericano llegaron a la conclusión de que la relación entre las actitudes y la conducta no sólo era importante para estudiar la política general del ejército, sino también en el estudio de casos específicos, tales como la moral de la tropa y las conductas manifestadas por los soldados en el frente de batalla. Esas primeras investigaciones produjeron un elevado número de estudios que hoy día son clásicos en este género (ver, por ejemplo, Hovland, Lumsdaine y Sheffield, 1949; Stouffer, Suchman, Devinney, Star y Williams, 1949, citados por Summers, 1976).

### **Conceptualización de las Actitudes**

Los investigadores han definido el constructo actitud de muchas maneras. Allport (1935) reportó más de 100 definiciones de este constructo y tal proliferación de conceptos no ha disminuido a través del tiempo. Algunas de estas definiciones señalan que “las actitudes constituyen una disposición fundamental que interviene junto con otras influencias en la determinación de una diversidad de conductas hacia un objeto o clase de objeto, las cuales incluyen declaraciones de creencias y sentimientos acerca del objeto y acciones de aproximación-evitación con respecto a él” (Cook y Selltitz, 1964, p. 10).

Por otra parte, Kidder y Campbell (1976) consideran que algunos términos aparentemente no relacionados, tales como: pulsión adquirida, creencia, reflejo condicionado, fijación, juicio, estereotipo, valencia y otros, son sinónimos funcionales del concepto de actitud. Ellos consideran que

todos estos términos describen residuos de experiencias pasadas que constituyen el material del que están hechas las actitudes.

Otra definición sobre actitud ha sido sugerida por Thurstone, quien la define como “la suma total de inclinaciones y sentimientos humanos, prejuicios o distorsiones, nociones preconcebidas, ideas, temores y convicciones acerca de un asunto determinado” (1928, p. 50; citado por Summers, 1976). También Cardno (1955) ha propuesto una definición de este constructo en los términos siguientes: “la actitud es la predisposición del sujeto para responder ante objetos sociales que en interacción con ciertas variables situacionales y posicionales, guían la conducta externa del sujeto” (p. 75).

Otros investigadores, como McGuinnis (1970) consideran que “las actitudes son disposiciones de respuestas similares a cualquier otra disposición que sirva de base a una ejecución aprendida” (citado por Wittaker, 1979, p. 239).

Aun cuando los ejemplos anteriores no pueden ser considerados como una muestra representativa del universo de definiciones sobre actitud, los mismos sí dan una idea de la diversidad de posiciones que los diferentes investigadores que se han ocupado del tema, tienen al respecto. Sin embargo, a pesar de estas diferentes interpretaciones sobre el significado de actitud, Summers (1976) ha identificado algunas áreas de coincidencia sobre la misma, entre las cuales están las siguientes:

1. Existe el consenso general de que una actitud es una predisposición a responder a un objeto y no la conducta efectiva hacia él.
2. La actitud es persistente, aunque no inmutable. Los diferentes estudios sobre las actitudes indican claramente que, aunque susceptibles

de cambio, la alteración de las actitudes requiere de un gran esfuerzo, tiempo y dedicación.

3. La actitud produce consistencia en las manifestaciones conductuales. Es decir, la actitud como variable latente da origen a la consistencia dentro de sus diferentes manifestaciones, que pueden tomar la forma ya sea de verbalizaciones hacia el objeto, expresiones de sentimiento acerca del objeto o de aproximación o evitación del objeto. Al respecto, Campbell (1963) ha señalado que “una actitud social del individuo es un síndrome de consistencia de respuestas hacia objetos sociales” (p. 96).

4. La actitud tiene una cualidad direccional; es decir, posee una dirección motivacional. En tal sentido, existe consenso, entre los investigadores, de que la actitud connota preferencia con respecto a resultados que implican al objeto, con base en las evaluaciones del mismo y en las impresiones positivas o negativas provocadas por aquél.

No obstante lo anterior, pareciera que la concepción que más predomina hoy día acerca de la actitud es la formulada por Katz y Stotland (1959) y por Krech, Crutchfield y Balachey (1962). De acuerdo con estos autores una actitud consta de los siguientes componentes: cognoscitivo, afectivo y conductual. El componente cognoscitivo incluye todas las creencias que se tienen acerca de un objeto, con base en el conocimiento que se tenga del mismo. Sin embargo, las creencias evaluativas parecen ser más importantes para la actitud como concepto de disposición. Estas abarcan las creencias acerca de las cualidades deseables e indeseables, aceptables o inaceptable, y buenas y malas. Krech y colaboradores (1962) consideran que las creencias acerca de cómo se debe tratar al objeto son también aspectos del componente cognoscitivo.

El componente afectivo es el aspecto de mayor importancia en una actitud (Fishbein, 1965). Por lo general, se le toma como respuesta emocional que va asociada con una categoría cognoscitiva a un objeto de la actitud. Este componente se forma a través de los contactos que hayan ido ocurriendo entre la categoría y ciertas circunstancias placenteras o desagradables. Es decir, para muchos investigadores el componente afectivo surge a través de una serie de procesos de asociación o condicionamiento entre los estímulos y algunos efectos recompensantes o de castigo. A partir de allí se generalizará la emoción asociada con la categoría a todos los reactivos que concuerdan con esta última.

Finalmente, el componente conductual de una actitud incluye el acto o la conducta que probablemente exhibirá un individuo en presencia de ciertos estímulos. Este componente permite anticipar la conducta que mostrará un individuo cuando se enfrenta con el objeto de la actitud.

En general, pareciera haber consenso entre los investigadores en cuanto a la relación que existe entre el componente cognoscitivo, particularmente las creencias evaluativas, y la disposición a responder al objeto. Igualmente pareciera haber acuerdo en cuanto a la compatibilidad direccional que existe entre los componentes cognoscitivos y conductual. En consecuencia si un docente cree que el Curriculum Nacional de Educación Básica ha sido bien concebido y que el mismo es útil para la formación del hombre venezolano del futuro, probablemente se incorpore activamente a participar en la administración de dicho currículo. También los investigadores parecieran estar de acuerdo en que existen lazos entre los componentes afectivo y conductual o de tendencia a la acción. De tal manera que si un individuo tiene un sentimiento negativo o desfavorable hacia un objeto, probablemente lo manifestará en su conducta al enfrentarse a dicho objeto.

## Medición de las Actitudes

La psicología, a través de su historia, ha tenido que enfrentar y superar muchos problemas relacionados con el logro de su estatus definitivo como disciplina de indagación científica. Uno de tales problemas ha sido el de descubrir y desarrollar métodos objetivos que permitan la comprensión, explicación predicción y control de diferentes clases de fenómenos relacionados con la conducta humana.

Como resultado, se ha logrado establecer un conjunto de constructos y de técnicas cuantitativas y cualitativas que han permitido a la psicología un enfoque científico de su objeto de estudio. Uno de los constructos formulados por la psicología es el de la actitud, el cual ya fue definido en la sección anterior.

Summers (1976) ha sugerido un procedimiento para medir los constructos psicológicos, en general, y de las actitudes, en particular. Este procedimiento incluye tres subprocesos, a saber: (a) la identificación de las muestras conductuales que se aceptan como base para hacer inferencias acerca del concepto (constructo) subyacente; (b) la recolección de las muestras conductuales; y (c) el tratamiento de las muestras de conducta para convertirla en una variable cuantitativa. A continuación se describen cada uno de estos subprocesos.

### **Identificación de Muestras Conductuales Aceptables**

Para realizar inferencias válidas de las actitudes de un grupo de individuos con respecto a un objeto, se requiere de la selección de muestras representativas de conducta que reflejen las creencias, sentimientos y tendencia a la acción de dichos individuos con respecto al objeto en cuestión.

Dada la correspondencia existente entre los diferentes componentes de la actitud, la inferencia puede ser realizada tanto a partir del componente cognoscitivo (creencias), como del afectivo (sentimientos) o conductual (tendencia a la acción).

La inferencia acerca de una actitud determinada puede ser realizada desde diferentes perspectivas, algunas de ellas son: (a) los informes personales sobre creencias, sentimientos y conductas; (b) la observación directa de la conducta manifiesta; (c) la reacción a estímulos parcialmente estructurados o interpretación de ellos, cuando implican al objeto de la actitud; (d) las tareas objetivas; y (e) las reacciones fisiológicas al objeto de la actitud o a representaciones de él (Cook y Selltiz, 1964).

Históricamente la medición de las actitudes ha descansado fundamentalmente sobre los “informes de sí mismo” (o informes personales) acerca de las creencias, sentimientos y conductas del individuo hacia el objeto de la actitud. Esta tendencia en la medición de las actitudes ha sido criticada por diferentes actores (Cook y Selltiz, 1964; Kazt y Scotland, 1976), quienes consideran que tal enfoque hace caso omiso de las múltiples características de las manifestaciones de las actitudes de un individuo en su conducta. Además, consideran que tratar de medir las actitudes a partir de una única base de inferencia sacrifica la oportunidad de examinar el postulado teórico de consistencia entre las diferentes manifestaciones conductuales de la actitud como concepto fundamental o variable latente. En este mismo sentido Webb y Salancik (1976) y Campbell y Fiske (1959) han planteado la conveniencia de establecer la relación empírica de la medición de la actitud con otro constructo asociado a fin de verificar su validez convergente; es decir, la confirmación del constructo por diferentes procedimientos de medición.

## **Recolección de Muestras Conductuales**

Los procedimientos de selección de muestras pueden ser estudiados tomando en cuenta la tipología de las bases de inferencia sugerida por Cook y Sellitz (1964), como criterio de organización. En tal sentido, se explican a continuación los diferentes procedimientos de recolección de muestras, de acuerdo con esta clasificación.

***Informe acerca de sí mismo.*** Como su nombre lo indica, este procedimiento consiste en preguntar directamente a la muestra de sujetos **qué siente** acerca de una persona, cosa o evento, **qué piensa** sobre tal individuo o qué estaría dispuesto a hacer por él. Este procedimiento toma en cuenta dos dimensiones primarias: (a) el método de aplicación (colectivo versus individual); y (b) el método de descripción (sujeto versus investigador). Al calcular todas las posibles permutaciones de estas dos dimensiones se obtienen cuatro procedimientos de recolección de muestras (Summers y Hammonds, 1969). Estos procedimientos son: (a) aplicación colectiva con descripción del sujeto; (b) aplicación en grupo con descripción del investigador; (c) aplicación individual con descripción del sujeto; y (d) aplicación individual con descripción del investigador.

***Observación de la conducta manifiesta.*** En este procedimiento el investigador no pregunta al sujeto acerca de sus creencias o sentimientos con respecto a un objeto, sino que observa directamente el comportamiento y a partir de ello infiere cual es la actitud del sujeto en cuestión. Este procedimiento tiene gran importancia por su contribución a la validación convergente del constructo actitud, puesto que la conducta manifiesta de una persona, en un momento dado, es el resultado del campo total de fuerzas que inciden sobre ella y no sólo dependen de las influencias que provienen de la actitud en estudio.

Existen varias formas de observación directa, las cuales varían de acuerdo con el momento de la observación. En algunos casos, es posible observar las reacciones de la persona al objeto de la actitud en situaciones de la vida real, sin que aquella se de cuenta que su conducta está siendo observada y registrada. En otras oportunidades, las observaciones directas de la conducta se realizan de manera planeada.

***Reacción a estímulos estructurados parcialmente.*** Este procedimiento consiste en controlar, hasta donde sea posible, la influencia de variables extrañas que influyen en la conducta, a fin de determinar con precisión la contribución de la actitud a esa conducta. Cuando se incluye al objeto dentro de un campo de estímulos muy estructurados, la respuesta de la persona al objeto tiene que estar determinada, en gran parte, si no totalmente, por la propia persona.

Generalmente, las técnicas para recoger muestras conductuales que son reacciones a estímulos estructurados parcialmente, y que comprometen al objeto de la actitud o a su representación simbólica, son similares a las técnicas de “informe sobre sí mismo”; sin embargo, existen algunas diferencias, entre las cuales están:

1. La intención encubierta del investigador que está presente en el procedimiento de reacción a estímulos estructurados parcialmente. En este caso el investigador utiliza un método indirecto, ya que el sujeto tiene presente, mientras contesta, categorías y dimensiones diferentes a las que el investigador emplea para interpretar la conducta del sujeto.

2. En este procedimiento se hace creer al sujeto que no hay respuestas correctas y que no hay criterio externo de corrección para evaluar la respuesta.

3. El sujeto debe elaborar su respuesta a la situación estímulo sin

ayuda del investigador. Se utiliza un estilo de respuesta libre, en el cual el investigador no limita o extiende artificialmente la amplitud de las posibles respuestas del sujeto. Por otra parte, este estilo de respuesta libre tan poco corre el riesgo de que el investigador imponga su propio conjunto de categorías y conceptos al sujeto.

4. Las instrucciones dadas por el investigador al sujeto en las colecciones de muestras que emplean estímulos estructurados parcialmente, son también parte integral de la situación estímulo, a la que reacciona el sujeto (ejemplo, láminas inestructuradas, manchas de tinta, dibujos sin palabras).

**Tareas objetivas.** En este procedimiento el sujeto es expuesto a una tarea concreta frente a la cual debe reaccionar. En esta situación el sujeto piensa que está describiendo al mundo tal como es y al hacerlo refleja, sin querer y sin saberlo, su propia interpretación de la realidad (Kidder y Campbell, 1976). Este procedimiento, al igual que los otros antes mencionados, también utiliza un método indirecto para inferir la actitud a partir de la conducta; sin embargo, en este caso se hace ver a la persona que existen respuestas correctas, las cuales deben ser identificadas al seleccionar las diferentes alternativas.

**Reacción fisiológica.** Las respuestas fisiológicas a la presencia de un objeto de actitud o a su representación pueden estar condicionadas o incondicionadas. El proceso de inferencia inherente al uso de respuestas fisiológicas condicionadas descansa en la teoría del aprendizaje. La mayoría de las respuestas fisiológicas son susceptibles de condicionamiento (Volkova, 1953), de allí que cuando una respuesta fisiológica se ha condicionado a un concepto evaluativo como bueno, malo, agradable, desagradable, la

presentación de un objeto que se percibe con la misma cualidad producirá la respuesta fisiológica.

En el caso de la respuesta incondicionada se supone que la magnitud de las respuestas fisiológicas está relacionada directamente con la convicción, el componente emocional de la actitud o con ambas cosas y que la selección debe mantenerse, dentro de ciertos límites en todas las personas (Cook y Selltiz, 1964).

### **Tratamiento de Muestras Conductuales**

Una vez identificadas y seleccionadas las muestras de interés, el investigador debe proceder a darle el tratamiento adecuado que permita determinar las dimensiones del conjunto. El procedimiento, por excelencia, utilizado para tratar las muestras conductuales consiste en la utilización de escalas numéricas. Sin embargo, para alcanzar la óptima eficacia en la manipulación numérica deben existir ciertas relaciones específicas entre la variable de estudio (actitud), las muestras (informes sobre sí mismo, observaciones directas, etc) y la escala numérica.

Una escala, como un instrumento de medición que es, puede ser definida como un conjunto de símbolos o números contruidos de tal manera que pueden ser asignados, de acuerdo con ciertas reglas, a los individuos (o a su conducta), a quienes se aplica la escala, en donde la asignación indica el grado en que el individuo posee el rasgo objeto de medición (Kerlinger y Lee, 2002). Puesto que una actitud es una variable hipotética o latente, más que una magnitud directamente observable, la medición de las actitudes consiste en la evaluación de las respuestas de un individuo que ha sido expuesto a un conjunto de situaciones estímulo (Green, 1954).

Este conjunto de situaciones es usualmente una muestra representativa de reactivos o ítemes acerca del objeto de la actitud frente a los que el

individuo responde con base en un conjunto de categorías específicas (por ejemplo, “de acuerdo” y “en desacuerdo”). El valor asignado a la respuesta de un individuo a un ítem determinado se denomina puntaje del ítem y el número resultante de la conjunción de los puntajes de los diferentes ítems de la escala representa la posición del individuo en el constructo objeto de estudio. De acuerdo con lo anterior, una escala de actitud puede ser definida como “un conjunto de ítems con sus respectivos puntajes” (Shaw y Wright, 1967, p. 15); o como ha sido establecido por Green (1954), cuando señala que “una escala de actitud es un conjunto de puntajes obtenidos de un conjunto de ítems” (p. 85).

## Métodos de Construcción de Escalas de Actitudes

Los procedimientos utilizados en la construcción de escalas de actitudes constituyen ampliaciones particulares del proceso general de elaboración de escalas psicológicas que, a su vez, se propone formular escalas de medidas para propiedades abstractas de la experiencia humana. En todo caso, cualquier procedimiento para elaborar escalas abarca tres grupos de variables, como son: los **estímulos**, un conjunto de objetos que hemos escogido (ejemplo, ítems de la escala); los **sujetos**, a quienes se presentan los estímulos; y las **respuestas**, que requiere la naturaleza de la situación experimental (Garner y Creelman, 1967).

El atributo que se pretende medir por escala es una propiedad de los sujetos. Para algunos atributos hay una dimensión física correspondiente a los estímulos, como en el caso en que el atributo es una experiencia sensorial de tamaño, brillantez o sonoridad. Para otros atributos, como las actitudes, no hay contraparte física en el estímulo. Esta circunstancia introduce una diferencia importante en la manera como se enfocan las mediciones en ambos tipos de atributos, lo que al mismo tiempo constituye una dificultad práctica en la construcción de las escalas de actitudes. Sin embargo, desde los trabajos pioneros de Thurstone y Chave en 1929 (citado por Edwards, 1957), ha habido un progreso importante en el desarrollo de técnicas de construcción de escalas de actitudes.

Entre los métodos reportados por la literatura para la medición de actitudes se encuentran los siguientes: (a) el método de intervalos iguales (Thurstone, 1929); (b) el método de los rangos sumados (Likert, 1932); (c) el método del escalograma (Guttman, 1944, 1947); (d) el método de

discriminación escalar (Edwards y Kilpatrick, 1948); y (e) la semántica diferencial (Osgood, Suci y Tannenbaum, 1957).

En esta obra, sólo se desarrollará el método de rangos sumados, por varias razones; (a) por ser uno de los métodos de construcción de escalas más usados en los ambientes académicos y profesionales; (b) porque con mucha frecuencia este método se usa de una manera inapropiada; y (c) porque el desarrollo de cada uno de los métodos de construcción de escalas de actitudes está fuera del alcance de esta obra. De todas maneras, el lector interesado en ampliar su información al respecto puede consultar las obras de Edwards (1957), Summers (1976), Dawes (1975), Manning y Rosenstock (1971), Henerson, Morris y Fitz-Gibbon (1978), entre otros.

### **El Método de Rangos Sumados**

Este método de construcción de escalas de actitudes fue desarrollado por Likert (1932), para lo cual siguió los pasos siguientes:

1. Seleccionó un conjunto de frases que representaban ítemes cuyos valores escalares, correspondientes a un determinado continuo psicológico, eran desconocidos.

2. Los ítemes fueron administrados a una muestra de  $n = 200$  sujetos, a quienes se les pidió clasificar dichos reactivos de acuerdo con el grado de favorabilidad que tenían hacia los mismos. Las respuestas fueron clasificadas con base en las siguientes cinco categorías: completamente de acuerdo, de acuerdo, indeciso, en desacuerdo y completamente en desacuerdo.

3. Se calculó para cada ítem la proporción de respuestas obtenidas en cada una de las cinco categorías, lo cual sirvió de base para la ponderación de cada una de ellas, encontrándose una serie de valores que se aproximan a la escala siguiente:

4	Completamente de acuerdo
3	De acuerdo
2	Indeciso
1	En desacuerdo
0	Completamente en desacuerdo

Sin embargo, se ha generalizado la práctica de utilizar valores comprendidos entre 5 y 1, en vez de entre 4 y 0, lo cual no altera la concepción de la escala. Dado que tales valores escalares fueron derivados empíricamente y han demostrado representar una buena aproximación para resolver el problema de la medición de las actitudes, quienes deseen usar este método de construcción de escalas no necesitan derivar nuevamente los pesos o valores escalares de las diferentes categorías, sino que simplemente procederían de la siguiente manera:

1. Se define el constructo u objeto psicológico a ser evaluado.
2. Se genera un conjunto de ítemes representativos del constructo, cuyo número debe ser aproximadamente tres veces la extensión final de la escala. Para tal fin, es indispensable elaborar un plan de operacionalización del constructo, el cual debe contener los elementos siguientes: la definición del propósito del instrumento, la conceptualización del constructo, la identificación de la dimensiones implícitas en el objeto de la medición, la formulación de los indicadores para cada dimensión y, finalmente, la redacción de los ítemes correspondientes a cada indicador. En relación con la redacción de los ítemes, Edwards (1957) ha formulado las recomendaciones siguientes:

- Evitar los reactivos que se refieren al pasado, en lugar del presente.

- No incluir reactivos que puedan ser interpretados como acciones de hecho.
- No usar reactivos que se presten a una interpretación ambigua.
- No incluir reactivos que sean irrelevantes al constructo.
- Desechar aquellos reactivos que pudieran ser aceptados favorablemente por todo el mundo o por casi nadie.
- Seleccionar los reactivos que cubran el dominio completo e contenido a que se refiere el constructo.
- Redactar los reactivos en forma sencilla, clara y directa.
- Los reactivos deben ser cortos, por lo general no deberían pasar de 20 palabras.
- Cada reactivo debería representar una sola idea o pensamiento.
- Evitar en los reactivos expresiones tales como: “todos”, “siempre”, “nadie”, “nunca”.
- Usar con precaución expresiones tales como: “solamente”, “meramente”, “justamente” y otras similares.
- Evitar el uso de vocablos o términos desconocidos para los sujetos que responden la escala.
- Se deben redactar igual número de ítems con sentido positivo y negativo hacia el objeto de la medición.
- Los ítems deben ser ordenados aleatoriamente y no siguiendo una determinada secuencia.

3. Se determina la validez de contenido de la escala, de acuerdo con el método de “jueces expertos”. Y, con base en ello, se hacen las revisiones que fueran necesarias a los reactivos.

4. Se selecciona una muestra representativa de sujetos, manteniendo como mínimo la proporción de dos sujetos por cada ítem de la escala en su versión original. Por lo general, se recomienda utilizar muestras que oscilan entre 180 y 200 sujetos.

5. Se administra la escala en su primera versión, a la muestra de sujetos seleccionada.

6. Se realiza el análisis de los ítems, tomando en cuenta los siguientes pasos:

- Se elabora una matriz de ítem por sujeto, con los resultados registrados en los diferentes formularios respondidos por los sujetos.
- Se suman los valores escalares de cada sujeto para obtener su puntuación total.
- Se ordenan jerárquicamente los sujetos de acuerdo con la puntuación obtenida por cada sujeto en el instrumento.
- Se forman grupos extremos (actitud más favorable versus actitud menos favorable) con base en los valores de la distribución de puntajes obtenidos por los sujetos de la muestra. Para tal fin, se calcula el 27 por ciento superior e inferior de la distribución. Por ejemplo, para una muestra de  $n = 200$ , el 27 por ciento serían 54 sujetos. En este caso se tomarían los 54 individuos con los puntajes más altos (grupo favorable) y los 54 con las puntuaciones más bajas (grupo desfavorable).
- Se calcula para cada ítem la media y la desviación estándar en ambos grupos (favorable y desfavorable).

- Se calcula la discriminación del ítem, mediante la prueba t de student, para muestras independientes, de igual tamaño. Al respecto se puede utilizar la siguiente fórmula:

$$t = \frac{M_{gf} - M_{gd}}{\sqrt{\frac{V_{gf}}{N_{gf}} + \frac{V_{gd}}{N_{gd}}}}$$

Donde:

$t$  = Es el estadístico, que representa el índice de discriminación del ítem;

$M_{gf}$  = Media del ítem en el grupo favorable;

$M_{gd}$  = Media del ítem en el grupo desfavorable;

$V_{gf}$  = varianza del ítem en el grupo favorable;

$V_{gd}$  = varianza del ítem en el grupo desfavorable;

$N_{gf}$  = número de sujetos en el grupo favorable;

$N_{gd}$  = número de sujetos en el grupo desfavorable.

- Para que un ítem sea aceptado, se requiere que el valor de la  $t$  observada sea mayor que el valor de la  $t$  teórica (valor de la tabla de distribuciones de  $t$  correspondiente, de acuerdo con los grados de libertad implícitos y un nivel de significación estadística mínimo del 5 por ciento ( $\text{Alpha} = 0,05$ ). En el caso de una muestra de  $n = 200$  y grupos (favorable y desfavorable) del tamaño  $n_1 = n_2 = 54$ , se requeriría que el valor de la  $t$  observada o calculada para cada ítem fuera mayor que 1,96 ( $t$  teórica).

Otros procedimientos que pueden ser utilizados para el análisis y selección de los ítems son: (a) la correlación ítem-total corregida y (b) la correlación ítem-factor en un análisis factorial. En el primer caso se considera que un coeficiente de correlación igual o mayor que 0,41 puede ser adecuado como criterio de retención del ítem en la escala. Este mismo criterio podría

ser aplicado en el caso de la correlación entre cada ítem y el o los factores asociados.

Con la aplicación del método de rangos sumados, lo que se busca es identificar entre 20 y 25 ítems que discriminen entre los sujetos que tienen alta puntuación en la escala (actitud favorable) y aquellos cuyo puntaje es bajo (actitud desfavorable).

Una vez aplicada la técnica estadística apropiada (prueba t, correlación ítem-total, correlación ítem-factor), para el análisis de los ítems, éstos se ordenan de acuerdo con los valores estadísticos obtenidos y se seleccionan los 20 o 25 reactivos que tengan los coeficientes más altos. Al respecto, es importante que la selección final de los ítems, se mantenga un balance apropiado entre ítems positivos y negativos. Por ejemplo, en una escala cuya versión final sea de 20 ítems, se espera que 10 de ellos estarán formulados positivamente; mientras que los otros 10 estarán fraseados en forma negativa.

7. Se realiza el estudio técnico de la escala, lo cual implica: (a) estimar la confiabilidad (consistencia interna, test-retest, equivalencia); (b) estimar la validez (constructiva, predictiva); (c) establecer las normas y las condiciones de estandarización para la aplicación e interpretación de los resultados del instrumento.

8. Se interpretan los resultados, lo cual se puede hacer de varias maneras:

- Comparando el puntaje de un sujeto con respecto a la norma establecida, en el grupo normativo o grupo de estandarización. Estas normas pueden ser con base en percentiles o en escalas estandarizadas (z, T, estantina). Cuando este tipo de instrumento se utiliza con un propósito de selección, el criterio mínimo que

normalmente se adopta es el del percentil 50 (la mediana) o la media, en el caso de una escala estandarizada. Se considera que una persona tendrá una actitud más favorable hacia un objeto psicológico determinado, en la medida que la puntuación del sujeto excede los valores normativos de tendencia central, particularmente la mediana y la media.

- Comparando la puntuación promedio del grupo donde ha sido aplicada la escala, con respecto a las normas del instrumento.
- Clasificando directamente los sujetos de acuerdo con las diferentes categorías de la escala. Así se podría saber, por ejemplo, cuantos sujetos tuvieron una actitud favorable (valores entre 4 y 5), desfavorable (valores entre 1 y 2) o neutra (valor igual a 3) hacia el objeto de estudio.

En general, los puntajes obtenidos en este tipo de escala permiten identificar la rata de cambio promedio obtenido en actitud hacia un objeto dado, como consecuencia de la manipulación de una variable experimental. Igualmente, tales puntajes pueden servir para comparar las calificaciones promedio en actitud de dos o más grupos. Finalmente, estos puntajes pueden ser correlacionados con los de otras escalas de actitudes o con otras medidas de interés.

## **Escala de Actitud Hacia la Innovación Educativa**

Con el propósito de ilustrar la aplicación del método de rangos sumados (llamado también método Likert) en la construcción de escalas de actitudes, se describe a continuación todo el proceso de construcción de la escala de actitud hacia la innovación educativa (ESCAHINE), desarrollada por Ruiz (1991). En tal sentido, se seguirá, como organizador, el esquema presentado en el capítulo 1 de esta obra para la construcción de instrumentos de medición en ciencias de la conducta. Dicho esquema incluye las fases siguientes: (a) determinar el propósito del instrumento, (b) decidir sobre el tipo de instrumento; (c) conceptualizar el constructo; (d) operacionalizar el constructo; (e) integrar el instrumento; (f) realizar la prueba piloto; y (g) realizar el estudio técnico.

### **Definición del Propósito del Instrumento**

Estudios recientes sobre el sistema educativo venezolano han puesto de manifiesto los nudos críticos que lo caracterizan y, a su vez, lo limitan para cumplir con los fines para los cuales fue creado (ver, por ejemplo, Banco Mundial, 1993; Consejo Nacional de Educación, 1993 y Comisión Presidencial para la Reforma del Estado, 1990). Al mismo tiempo, pareciera haber consenso entre los investigadores, políticos, empresarios, educadores y padres o representantes, acerca de la necesidad de actuar de manera distinta frente a los problemas de la educación; es decir, se piensa que la innovación puede ser una alternativa válida frente a la crisis del sistema educativo.

Algunos autores han coincidido en su apreciación sobre la innovación educativa como un proceso deliberado dirigido a producir cambio social (ver Havelock y Huberman, 1980, Olivares, 1984; Albornoz, 1987). Es decir, lo

que caracteriza y justifica una innovación educativa es la posibilidad de ofrecer soluciones creativas a problemas viejos no resueltos, o a problemas nuevos para los cuales todavía no se conocen cuáles son las soluciones más apropiadas.

Las investigaciones realizadas, a nivel nacional e internacional, en relación con los procesos de innovación educativa han permitido identificar algunos obstáculos que se oponen a la realización exitosa de los proyectos de innovación. Algunos de esos obstáculos son: (a) la ingerencia político-partidista en la toma de decisiones técnicas sobre la conducción de los proyectos; (b) falta de recursos humanos capacitados; (c) las limitaciones económicas de los países; y (d) la resistencia al cambio por parte del personal docente.

En relación con el último de los obstáculos señalados, CINTERPLAN (1987) ha planteado lo siguiente:

*En la intensa actividad institucional que ha desarrollado CINTERPLAN entre los países del continente, en materia de innovación, ha quedado en evidencia que el docente es -entre otros múltiples- una de las principales limitaciones que encuentran los esfuerzos por difundir, impulsar y viabilizar cambios significativos en la educación ... Ha sido el profesor -que debe interpretar los nuevos conceptos y que debe manipular profesionalmente los nuevos recursos- quien se manifiesta más reacio e indiferente a las innovaciones, cuando no se presenta definitivamente contrario a cualquier cambio que altere su concepción educativa y su quehacer pedagógico cotidiano (p. 15).*

Esta impresión que se tiene sobre la actitud del docente hacia el cambio educativo ha sido compartida también por otros autores como, por ejemplo,

Ministerio de Educación (1970); Olivares, 1984; Albornoz, (1987); y Ruiz (1991). Esta limitante es crucial, si se toma en cuenta que cualquier programa o proyecto de innovación educativa que se adopte con un propósito definido, afectará de alguna manera la actividad del docente. De allí la necesidad de poder contar con su receptividad, cooperación, motivación y, en última instancia, con su actitud favorable, a fin de garantizar el éxito de la acción innovadora. Es decir, el mejor proyecto de innovación educativa será expuesto al fracaso, si no cuenta con el apoyo suficiente de los docentes, quienes deberán implementarlo y administrarlo académicamente a nivel operativo.

En todo proceso de innovación educativa, hay que tomar en cuenta la existencia del fenómeno denominado “resistencia al cambio”, el cual puede tener una motivación racional o irracional (Barnes, 1967). En el primer caso, la resistencia al cambio puede estar fundada en sentimientos emocionales que conducen a una posición tradicionalista frente a la institución escolar. También puede estar basada en el “temor a lo desconocido”, si no ha habido suficiente información previa, o en la amenaza cierta de lo que, para algunas posiciones o estatus, pudiera significar la adopción de un determinado programa de innovación.

La manera irracional de resistirse al cambio, puede adoptar varias formas. Una de ellas se expresa en una actitud conservadora que se manifiesta en una conducta negativa frente a todo intento de cambio. Otras razones para resistirse al cambio son: la inercia o hábito en la manera de realizar el trabajo diario y la baja expectativa que el docente tiene frente a la posibilidad de éxito de cualquier programa de innovación en su institución.

Lo anterior justifica el diseño de una estrategia apropiada que permita vencer, de una manera total o parcial, la natural resistencia al cambio que

comporta toda acción innovadora. Una de esas estrategias consiste en modificar las actitudes de aquellos docentes que son menos favorables al cambio en educación.

Pero para ello es necesario conocer, previo al inicio de cualquier proyecto de innovación, cuál es la actitud que tienen los docentes en relación con un determinado proyecto. Esto permitirá anticipar cuál es la magnitud de la inversión que es necesario realizar en esfuerzo, tiempo y recursos materiales, para alcanzar las mejores condiciones posibles en cuanto a la aceptación y cooperación por parte de los docentes, a fin de llevar adelante, con éxito, un proyecto de innovación educativa.

Lo anterior significa que es necesario poder contar con un instrumento de medición que permita discriminar entre los docentes creativos e innovadores y aquellos que no lo son o lo son en menor grado, para determinar quienes requerirán mayor o menor intervención a fin de mejorar su actitud hacia el proyecto en cuestión y así lograr su participación activa y comprometida en el mismo. De allí que el presente estudio tenga como propósito principal desarrollar un instrumento que permita evaluar las actitudes de los docentes hacia la innovación educativa.

En otras palabras, con el desarrollo de la escala de actitud hacia la innovación educativa se intentaba identificar en qué medida los docentes responsables de llevar a efecto un programa o proyecto de innovación educativa se sentían inclinados a participar conscientes, activa y solidariamente en el mismo, a objeto de lograr sus objetivos con alta calidad.

Se pensó que el tipo de instrumento más apropiado para cumplir con el propósito propuesto era la escala, entre otras razones, porque el constructo que se deseaba estudiar era la actitud, en cuyo campo existe ya una metodología específica para su medición.

## **Conceptualizar el Constructo**

Después de una amplia revisión de la literatura en relación con las actitudes y la innovación educativa, el constructo **Actitud hacia la Innovación Educativa** quedó definido como sigue: la predisposición del docente para reaccionar y actuar, favorable o desfavorablemente, en situaciones educativas que suponen cambios deliberados, dirigidos a modificar cualitativamente uno o más de los componentes del sistema educativo o de instrucción, de acuerdo con el conocimiento y sentimiento del sujeto hacia el objeto específico de la actitud.

De acuerdo con la literatura, se asumió que el constructo tenía cuatro dimensiones. En este sentido, la actitud hacia la innovación sería enfocada desde el punto de vista de los siguientes aspectos del sistema educativo:

1. Los **objetivos y funciones**, los cuales se refieren a aquellas innovaciones dirigidas a cambiar los objetivos y funciones del sistema educativo tanto a nivel global como del plantel.

2. La **administración**, incluye todas aquellas innovaciones relacionadas con los procesos administrativos, tales como: estructura organizacional, gerencia, logística y supervisión.

3. El desempeño de roles y las relaciones interpersonales, comprende las innovaciones que están dirigidas a modificar los roles que desempeñan diferentes funcionarios que operan el sistema educativo (supervisores, directores, coordinadores, docentes especialistas, docentes de aula, etc).

4. El **currículum**, incluye las innovaciones que afectan los diferentes componentes curriculares: los objetivos, los contenidos, los métodos instruccionales, los recursos para el aprendizaje y la evaluación.

Con base en la conceptualización anterior, se plantearon las siguientes

interrogantes:

1. ¿Cuáles eran la validez de contenido y constructiva de la ESCAHINE?
2. ¿Cuál era la confiabilidad de consistencia interna de la ESCAHINE?
3. ¿Cuáles eran las normas para la administración e interpretación de los resultados de la ESCAHINE?
4. ¿Existían diferencias significativas en la ejecución en la escala entre las submuestras de docentes en servicio y de docentes en formación?
5. ¿Cuál era la relación entre la actitud hacia la innovación y variables tales como: motivación al logro y autoritarismo?
6. ¿Cuál era la relación entre la actitud hacia la innovación y la edad de los sujetos?
7. ¿Existían diferencias significativas en la ejecución de los sujetos en la ESCAHINE en cuanto al sexo?

### **Operacionalización e Integración**

Una vez definidas las dimensiones del constructo, se identificaron los diferentes indicadores por dimensión, como se observa en el 14. Finalmente, se formularon los ítems correspondientes a la primera versión de la escala, la cual quedó integrada por 80 reactivos (20 por cada dimensión), presentados y fraseados en un formato tipo Likert.

Con los 80 reactivos formulados para la primera versión de la escala se procedió a establecer su validez de contenido, a través del procedimiento de jueces expertos, los cuales estuvieron integrados por dos psicólogos sociales, dos educadores y un investigador educacional. Como resultado de esta primera revisión del instrumento, fue necesario revisar y modificar 15 ítems

del mismo, los cuales fueron integrados nuevamente a la escala para llevar a efecto su estudio piloto.

*Cuadro 14. Dimensiones e Indicadores de la ESCAHINE*

Dimensiones	Indicadores
Objetivos y Funciones	Desarrollo de la personalidad Formar un hombre sano, culto y democrático Valorización de la familia Valorización del trabajo Participación activa, consciente y solidaria Identidad nacional Conciencia conservacionista Valoración del arte Formación de valores éticos Desarrollo del pensamiento reflexivo, crítico y creativo
Administración	Organización del Sistema Educativo Investigación educacional Planificación educacional Recursos financieros Recursos humanos Supervisión educacional Descentralización Toma de decisiones Planta física y dotación Orientación educacional
Desempeño de Roles y Relaciones Interpersonales	Comunicación Estilo de liderazgo Trabajo en equipo Rol del personal directivo Rol del personal docente Rol del alumno Rol de los padres Organización del plantel Normas y procedimiento Evaluación del desempeño
Currículo	Objetivos instruccionales Contenidos Procesos cognitivos Métodos de enseñanza Recursos para el aprendizaje Evaluación Microplanificación

## **Prueba Piloto**

La prueba piloto de la primera versión de la ESCAHINE fue realizada en una muestra intencional de  $n = 256$  sujetos, de los cuales 156 (61 por ciento) eran docentes en formación, mientras que 100 (39 por ciento) eran docentes en servicio. Los docentes en formación eran cursantes de los últimos dos semestres en diferentes Institutos de la Universidad Pedagógica Experimental Libertador (UPEL); mientras que el resto eran docentes de distintos planteles públicos y privados, de Educación Básica y de Media, Diversificada y Profesional, de la Zona Metropolitana de Caracas.

Los criterios utilizados para la selección de la muestra fueron los siguientes:

1. El tamaño de la muestra se estimó utilizando como criterio una relación de tres sujetos por ítem, lo cual en este caso suponía la utilización de una muestra mínima de  $n = 240$  sujetos (resultado de multiplicar 80 reactivos de la escala por 3 sujetos). Este criterio de tres sujetos por ítem representa un punto intermedio entre los valores 5 a 2 sujetos por ítem recomendados por algunos especialistas (ver Nunnally, 1967; Guertin y Bailey, 1970), como el número mínimo aceptable de sujetos por ítem para el desarrollo de una escala a través de la técnica del análisis factorial.

2. El segundo criterio consistió en la selección de sujetos que representaban un amplio espectro del constructo objeto de estudio. De allí que se seleccionaran individuos heterogéneos en cuanto a las variables: edad, escolaridad, sexo, dependencia del plantel y estatus laboral. Tal decisión se tomó en virtud de que los estudios anteriores realizados con muestras venezolanas y extranjeras han demostrado que los docentes

tienden a ser conservadores; es decir, propenden a mantener el “estatus quo” y, por consiguiente, son pocos proclives a involucrarse en actividades innovadoras, que supongan cambios importantes en su rutina diaria de trabajo (ver Albornoz, 1987, Olivares, 1984; Ministerio de Educación, 1970).

### **Estudio Técnico**

Esta etapa se realizó en cuatro fases: análisis de ítem, estimación de la validez, estudio de la confiabilidad, y estandarización del instrumento. A continuación se describen cada una de éstas.

**Análisis de ítem.** Los 80 ítems de la ESCAHINE se analizaron mediante tres procedimientos diferentes, los cuales fueron: el análisis factorial, el uso de la prueba t en grupos extremos y la correlación ítem-total corregida. En el primer caso, se corrieron varias soluciones factoriales hasta obtener una donde se pudo evidenciar que la estructura de los ítems reflejaba dos factores claramente definidos en un total de 32 reactivos, los cuales fueron seleccionados con base en el siguiente criterio: cada ítem debía tener una correlación en un factor igual o mayor que 0,36 y, al mismo tiempo, tener una correlación muy baja en los demás factores.

El segundo procedimiento utilizado consistió en determinar la discriminación de los ítems, mediante el uso de la prueba t con grupos extremos, el cual ya fue descrito anteriormente en este mismo capítulo. Finalmente, el tercer procedimiento consistió en establecer la correlación entre la distribución de cada ítem con la distribución total, una vez que se ha removido, de dicha distribución, los valores del ítem que se está correlacionando.

Los resultados de los ítems seleccionados mediante los tres métodos antes expuestos se presentan en el Cuadro 15. Como se puede observar, los

tres métodos dan resultados equivalentes.

**Validez de la ESCAHINE.** Como ya se mencionó, previo al estudio piloto, la escala fue sometida a un estudio de validez de contenido. Posteriormente, después de someter el instrumento al estudio piloto, se estableció la validez de constructo, lo cual se hizo, en primer lugar, mediante la técnica del análisis factorial. De esta manera se puso en evidencia que existen sólo dos dimensiones o factores representativos de la estructura subyacente en los ítemes del instrumento. En consecuencia, se pudo comprobar que, en lugar de las cuatro dimensiones independientes, formulada hipotéticamente (objetivos y funciones, administración, desempeño de roles y relaciones interpersonales, y curriculum) en la conceptualización de la escala, ésta realmente está formada sólo por dos dimensiones. Así, el primer factor, resultado del análisis factorial, integró las dimensiones: curriculum y, desempeño de roles y relaciones interpersonales; mientras que el segundo factor integró las dimensiones: administración y, objetivos y funciones. Estos factores han sido denominados tentativamente: (a) actitudes hacia la innovación en los aspectos “gerenciales” del sistema educativo; y (b) actitudes hacia la innovación en los aspectos “operacionales” del sistema educativo, respectivamente. La estructura factorial de la escala puede ser observada en el Cuadro 16.

La validez de constructo de la escala también fue confirmada a través del método sugerido por Campbell y Fiske (1959), como se planteó en el Capítulo 2 de esta obra. Estos autores, han planteado que la validez de constructo de un instrumento puede ser estimada mediante dos procedimientos: (a) el de la validez convergente; y (b) el de validez discriminante.

*Cuadro 15. Resultados del análisis de ítem de la ESCAHINE por tres*

*métodos diferentes*

Item	Análisis Factorial	Grupos Extremos (Prueba t)	Correlación Item-Total Corregida
2	.604	2,53	0,358
3	.667	2,05	0,300
9	.696	2,96	0,505
10	.723	2,83	0,470
11	.553	2,78	0,443
12	.821	2,10	0,309
13	.644	2,40	0,339
17	.666	2,62	0,348
27	.663	2,65	0,393
29	.806	2,55	0,369
30	.604	2,83	0,469
32	.660	2,70	0,409
34	.889	2,42	0,334
35	.666	2,79	0,429
38	.668	2,76	0,420
39	.648	2,97	0,509
41	.795	2,05	0,301
43	.940	2,53	0,362
45	.763	2,38	0,317
50	.867	2,75	0,420
53	.862	2,53	0,358
54	.920	2,74	0,417
55	.828	2,83	0,470
58	.784	2,30	0,430
67	.813	2,53	0,363
69	.809	2,80	0,440
72	.856	2,80	0,437
73	.685	3,25	0,567
74	.959	3,05	0,546
78	.794	2,58	0,374
79	.823	3,00	0,531
80	.793	3,05	0,545

*Cuadro 16. Cargas Factoriales de los 32 Items de la ESCAHINE*

Item	Factor 1	Factor 2
2	-.232	.604
3	-.221	.667
9	.156	.696
10	-.161	.723
11	-.231	.553
12	-.192	.821
13	-.159	.644
17	-.191	.666
27	-.230	.663
29	-.210	.806
30	-.244	.604
32	-.210	.660
34	-.251	.889
35	-.187	.666
38	-.221	.668
39	-.198	.648
41	.795	.192
43	.940	.183
45	.763	.193
50	.867	.214
53	.862	.229
54	.920	.213
55	.828	.213
58	.784	.194
67	.813	.152
69	.809	.183
72	.856	.171
73	.685	.092
74	.959	-.138
78	.794	-.151
79	.823	.097
80	.793	.006

En el primer caso, se trata de demostrar empíricamente que existe una relación hipotetizada entre el constructo objeto de estudio y otro que se utiliza como criterio. En el segundo caso, se intenta probar empíricamente que no existe asociación entre el constructo de interés y otro con el cual aquél no debería estar relacionado ni teórica ni lógicamente.

En la situación particular de la ESCAHINE, sólo se utilizó el procedimiento de validez convergente. En tal sentido se utilizaron dos medidas como criterio; ellas fueron: la escala de necesidad de logro de Merhabian (1969), versión de Salom de Bustamante (1981) y la escala de autoritarismo (escala F) de Adorno y colaboradores (1969). En el primer caso, se hipotetizó la existencia de una correlación moderada y positiva entre el nivel de motivación de logro y la actitud hacia la innovación educativa; mientras que en el segundo, se conjeturó la existencia de una correlación moderada y negativa entre los puntajes de la escala F y la ejecución de los sujetos en la ESCAHINE. Ambas hipótesis fueron confirmadas, como lo indican las coeficientes de correlación obtenidas, los cuales fueron de 0,55 y -0,74 (ver Cuadro 17).

*Cuadro 17. Matriz de intercorrelación entre la actitud hacia la innovación educativa, los factores de la escala y otras variables relacionadas*

Variables	2	3	4	5	6
1. Escahine	.76	.61	.55	-.74	-.45
2. Factor 1		.56	.38	-.53	-.41
3. Factor 2			.43	-.48	-.39
4. n-Logro				.25	.14
5. Autoritarismo					.30
6. Edad					

### **Confiabilidad de la ESCAHINE**

Se estimó la confiabilidad para la escala total y por submuestras. En tal sentido, se establecieron dos tipos de confiabilidad: (a) consistencia interna (métodos: Alpha de Cronbach y de Hoyt); (b) test-retest. Los resultados se presentan en el Cuadro . Como se puede observar, la confiabilidad de consistencia interna, para la escala total fue de 0,83 y 0,84, por los métodos Alpha de Cronbach y Hoyt; mientras que el índice de estabilidad (método de test-retest) fue de 0,82. Por otra parte, el índice de confiabilidad, por submuestras oscilan entre 0,80 y 0,95. Estos resultados indican que la confiabilidad de la ESCAHINE es aceptable en relación con los fines que se propone (ver Cuadro 18).

*Cuadro 18. Coeficientes de confiabilidad por diferentes métodos para la muestra total y por submuestras*

Muestras	n	Consistencia Interna		Estabilidad Test-retest
		Alpha	Hoyt	
Total	256	0,84	0,83	0,82
Docente en Servicio	100	0,82		0,81
Docente en Formación	156	0,80		0,80
Factor 1	256	0,95	0,95	
Factor 2	256	0,93	0,93	
Varones	80	0,81		0,80
Hembras	176	0,83		0,89

### **Estandarización de la ESCAHINE**

En esta sección se describen las normas obtenidas para el instrumento y las condiciones bajo las cuales el mismo debe ser administrado. Como ya se describió en la sección del estudio piloto, el grupo normativo estuvo formado por 256 sujetos, docentes en servicio y en formación, con una edad promedio de 29 años ( $s = 7,5$ ), pertenecientes a planteles públicos y privados, correspondientes a los niveles de Educación Básica y Media, Diversificada y Profesional. Se establecieron dos tipos de normas: (a) con base en percentiles y cuartiles; y (b) con base en la escala T. Esta escala tiene una media igual a 50 y una desviación estándar igual a 10. Para el cálculo de los valores de la distribución T se requiere previamente haber transformado los puntajes directos en valores de z. La fórmula utilizada es como sigue:  $T = z(10) + 50$ .

**Normas percentiles y cuartiles.** Se calcularon los rangos percentiles para las puntuaciones directas de la escala, las cuales estuvieron comprendidas entre los valores 87 y 146. Asimismo, se establecieron los valores cuartilares, los cuales fueron: 117, 123 y 128, correspondientes a Q1,

Q2 (o la Mediana) y Q3. Por lo tanto, una calificación igual o mayor a 123 es indicadora de una actitud favorable hacia la innovación educativa; mientras que una puntuación comprendida entre 32 y 122, refleja una actitud desfavorable o negativa hacia el constructo, siendo más desfavorable en la medida que se aproxima al límite inferior del intervalo mencionado (ver Cuadro 19)

***Normas en Puntajes de T.*** La estandarización de las calificaciones se hizo mediante la utilización de la escala T. Así, para cada puntaje directo (o puntaje bruto) se obtuvo una puntuación estandarizada, como se indica en el Cuadro. Como se puede observar, el puntaje que corresponde a la media estandarizada de 50, es el 121, el cual está dos puntos por debajo de la mediana. Esta pequeña diferencia se debe a que la distribución de las calificaciones no es completamente normal, sino ligeramente asimétrica hacia la izquierda. Sin embargo, dicha diferencia no afecta el criterio de decisión que permite determinar cuando la actitud es o no favorable, ya que la misma abarca un rango que iría, en los puntajes escalares, desde 3,5 hasta 4,49, los cuales equivaldrían al rango 112 – 144 en la distribución de puntajes brutos (ver Cuadro 20).

*Cuadro 19. Normas percentiles de la ESCAHINE*

<b>Puntaje: Directo</b>	<b>146–126 Percentil</b>	<b>Puntaje Directo</b>	<b>125–105 Percentil</b>	<b>Puntaje: Directo</b>	<b>104–87 Percentil</b>
146		125		104	
145	99	124	58	103	8
144		123	<b>Q<sub>2</sub></b>	102	
143		122		101	
142	98	121	44	100	6
141		120		99	
140		119		98	
139	97	118	30	97	4
138		117	<b>Q<sub>1</sub></b>	96	
137		116		95	
136	96	115	19	94	3
135		114		93	
134		113		92	
133	91	112	16	91	2
132		111		90	
131		110		89	
130	83	109	13	88	1
129		108		87	
128		107			
127	<b>Q<sub>3</sub></b>	106		9	
126	73	105			

*Cuadro 20. Puntajes directos y estandarizados de la ESCAHINE*

<b>Puntaje: Directo</b>	<b>146- 127 T</b>	<b>Puntaje: Directo</b>	<b>126-107 T</b>	<b>Puntaje: Directo</b>	<b>106-87 T</b>
146	67,87	126	53,57	106	39,28
145	67,14	125	52,85	105	38,57
144	66,42	124	52,14	104	37,86
143	65,70	123	51,53	103	37,14
142	65,00	122	50,71	102	36,42
141	66,28	121	50,00	101	35,70
140	63,57	120	49,28	100	35,00
139	62,86	119	48,57	99	34,28
138	62,14	118	47,85	98	33,57
137	61,43	117	47,14	97	32,86
136	60,71	116	46,43	96	32,14
135	60,00	115	45,71	95	31,42
134	59,28	114	45,00	94	30,71
133	58,57	113	44,28	93	30,00
132	57,86	112	43,57	92	29,28
131	57,14	111	42,86	91	28,57
130	56,42	110	42,14	90	27,85
129	55,71	109	41,42	89	27,14
128	54,11	108	41,43	88	26,42
127	54,28	107	40,00	87	25,71

## **Administración de la ESCAHINE**

La escala puede ser administrada en grupo o individualmente, siempre y cuando los sujetos tengan características similares a las del grupo normativo. Se aplica en 8 minutos y mantiene un balance apropiado en cuanto al número de ítems fraseados positiva y negativamente. Los reactivos negativos se identifican en la escala con un asterisco. Para obtener la puntuación total de cada sujeto, se suman los valores de las respuestas dadas a los ítems de la escala, siendo la puntuación máxima posible 160 y la mínima 32. Antes de sumar los valores escalares, es necesario revertir los valores de los ítems negativos; es decir, que si un ítem negativo fue puntuado, por ejemplo, con el valor 5 de la escala, al ser revertido, dicho valor se transforma en 1 y viceversa. Lo mismo ocurriría si el valor puntuado fuera 4, el cual pasaría a ser 2 y viceversa. No obstante, el valor 3 se mantiene en todos los casos.

***Desempeño de los sujetos en la ESCAHINE.*** La forma de la distribución de puntajes obtenida, de acuerdo con la ejecución de los sujetos de la muestra en la escala, tiene una forma ligeramente asimétrica hacia la izquierda (sesgo = -0,43). Ello indica que la mayoría de los sujetos obtuvo puntajes relativamente altos en el instrumento, los cuales se concentraron en la porción derecha de la distribución. La media aritmética y la desviación estándar obtenidas para el grupo total (N = 256) fue de 121 y 14 respectivamente; mientras que para las submuestras de docentes en servicio (n1 = 100), y docentes en formación (n2 = 156), estos estadísticos alcanzaron los valores de M = 119 y s = 16, para el primer caso; y de M = 125 y s = 12, para el segundo (ver Cuadro 21). Esta diferencia observada en la ejecución de estas dos submuestras fue estadísticamente significativa, como se evidencia mediante el valor de  $t(254) = 3,41$  ( $p < 0,01$ ), lo cual sugiere que

los docentes en formación tienen una actitud hacia la innovación educativa más favorable que los docentes en servicio (ver Cuadro 21)

*Cuadro 21. Medias, desviaciones estándares y prueba t por submuestras*

Muestras	N	Media	D.E.	t	p
Total	256	121	14		
Docentes en Servicio	100	119	16	3,41*	<0,01
Docentes en Formación	156	125	12		

Las submuestras de docentes en servicio y en formación también difieren significativamente en las variables: edad, necesidad de logro y autoritarismo (ver Cuadro 22).

*Cuadro 22. Medias, desviaciones estándares y valores de t en variables seleccionadas, por submuestras*

Variables	Docentes en Formación		Docentes en Servicio		t
	M	D.E	M	D.E.	
Edad	25	6,0	33	9,0	3,54
n-Logro	53	9,0	28	14,0	5,36
Autoritarismo	2,5	0,83	3,8	0,7	4,32

*Nota: Todos los valores de t fueron estadísticamente significativos, al menos, al nivel de 0,01.*

Finalmente, si se observa que la ejecución de los sujetos de acuerdo con el sexo, se puede apreciar que los varones (n = 80) obtuvieron una media aritmética en su ejecución de M = 123 (s = 16,5); mientras que en el grupo de

las hembras fue de  $M = 120$  ( $s = 12,8$ ). Sin embargo, esta diferencia no fue estadísticamente significativa, por lo cual con base en estos datos, no hay razón para pensar que los varones en general, tengan una actitud hacia la innovación educativa más favorable que las hembras o viceversa (ver Cuadro 23).

*Cuadro 23. Medias, Desviación Estándar y Prueba t por Sexo*

Muestras	N	Media	D.E	t
Total	256	121	14	
Varones	80	123	16,5	
Hembras	176	120	12,8	(ns)

# Pruebas de Rendimiento Académico

Desde el inicio mismo de la escuela, como institución destinada a administrar la educación formal, los docentes idearon diferentes maneras de apreciar los efectos de la enseñanza sobre el comportamiento del alumno. No podía ser de otra manera, por cuanto, los procedimientos empleados para estimar el grado de aprovechamiento de los conocimientos impartidos en el ámbito escolar o académico, constituyen parte del proceso de enseñanza-aprendizaje y están íntimamente vinculados con el sistema de calificación y de promoción de los alumnos.

Estos procedimientos evaluativos han variado, en su forma, rigurosidad y efectividad, a través del tiempo. Por ejemplo, en Venezuela hacia 1910, el procedimiento principal para apreciar el nivel de rendimiento del alumno era el juicio emitido por la junta de docentes evaluadores, basados exclusivamente en el examen oral (ver Mudarra, 1978).

Esta práctica evaluativa, que en sus comienzos se constituyó en una verdadera institución respetable, a la larga se prostituyó y degeneró en un verdadero adefesio, sin ninguna credibilidad. Al respecto Mudarra (1978) ha señalado que:

*Se llegó a tal corrupción en la comprobación de los conocimientos de los escolares, y hubo tal profusión de institutos que otorgaban certificados de instrucción primaria y hasta títulos de bachiller, que el Estado consideró indispensable su intervención en el régimen de exámenes (p. 106).*

Así, continúa Mudarra, por Ley de 1915, fueron separadas las funciones de docente de aula y de examinador. Se creó el Supremo Consejo Nacional de Educación, que tenía una función contralora de todo el sistema educativo, particularmente en lo concerniente a la función evaluadora. También se establecieron, además del examen oral, ya existente, las pruebas escritas y prácticas, de acuerdo con la naturaleza de la asignatura.

Estos procedimientos de evaluación del rendimiento estudiantil se han mantenido, a través del tiempo, como parte de la rutina del trabajo escolar, a pesar de los avances científicos logrados en el campo de la evaluación educacional y de las innovaciones incorporadas, al respecto, en la legislación vigente (ver Ministerio de Educación, 1980; 1986). Tal circunstancia desvirtúa los propósitos mismos de la evaluación de los aprendizajes e impide obtener una retroalimentación válida y confiable que permita fundamentar la toma de decisiones acerca de los cambios que se estimen necesarios hacer en los diferentes componentes del currículo, en función de las necesidades del alumno y de las exigencias del sistema social.

En el ámbito internacional, la tendencia hacia la apreciación objetiva del rendimiento escolar centrada en la medición, mediante el uso de pruebas estandarizadas, se desarrolló, particularmente en los Estados Unidos de Norte América, durante el primer tercio del presente siglo y obedeció a varias razones, entre las cuales están:

1. La consolidación del paradigma clásico de la ciencia, el cual se hizo extensivo hasta las ciencias sociales que, desde el punto de vista epistemológico, se fundamentaba en el positivismo lógico y metodológicamente en la experimentación y en la expresión cuantitativa de sus fenómenos.

2. La influencia del desarrollo de la ciencia de la Administración, en

el contexto de la transformación industrial de los Estados Unidos de Norte América. En este sentido, fueron particularmente determinantes, los trabajos de Taylor y Fayol en relación con la organización científica del trabajo, en la cual se privilegiaban los conceptos de eficiencia y medición, expresados a través de los procesos administrativos de previsión, organización, dirección, coordinación y control (ver Díaz- Barriga, 1987).

3. El auge de la teoría psicométrica, que sirvió de base para el desarrollo de instrumentos de medición estandarizados, que han sido utilizados no sólo en el campo de la psicología clínica, sino también en otras áreas de aplicación como la industria, las fuerzas armadas y, por supuesto, en el sistema educativo.

Con el transcurrir del tiempo, el énfasis puesto en la medición educacional fue tal que algunos docentes no-especialistas llegaron a confundirla con la evaluación misma. Hoy día, es bien sabido que los términos medición y **evaluación**, aun cuando están íntimamente relacionados en el contexto educacional, no son sinónimos. La **medición** se refiere a la cuantificación de los tipos de aprendizajes alcanzados por los estudiantes, los cuales se evidencian a través del uso de diferentes instrumentos; mientras que la evaluación implica, más bien, los juicios valorativos que el docente o el evaluador emite sobre los resultados del trabajo escolar, con base en un criterio determinado (objetivo o subjetivo). En este sentido, la evaluación es una función de naturaleza eminentemente cualitativa, lo cual no tiene por que ser necesariamente sesgada.

La “objetividad” en la evaluación depende de la naturaleza del criterio con base en el cual se formule el juicio valorativo. En ello influye, la posición epistemológica del docente o evaluador, la naturaleza del aspecto a evaluar (cognoscitivo, afectivo, psicomotor) y la objetividad, validez y confiabilidad

implícitas, en el procedimiento de registro de las evidencias de cambios conductuales ocurridos en el estudiante, en el lapso transcurrido entre el inicio y el final de una secuencia de aprendizaje (año escolar, semestre, trimestre).

### **Modelos de Evaluación**

Históricamente han existido diferentes teorías, modelos y enfoques de la evaluación (ver, por ejemplo, Tyler, 1942; Scriven, 1967; Stake, 1967; Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman y Provus, 1971; Provus, 1973; Popham, 1975, 1978; Glaser, 1963; Cronbach, 1963; Alkin, 1969). Guba y Lincoln (1982, 1992) han hablado de cuatro generaciones de evaluación, en el siguiente orden: como medición, descripción, juicio y negociación.

De lo anterior, se deduce lo difícil que resulta llegar a una posición de consenso acerca de qué, cómo, cuándo y para qué evaluar. La posición emergente, al respecto, enfatiza la evaluación cualitativa y naturalística, centrada en la negociación de los actores. En este sentido, la evaluación es vista como una construcción significativa del individuo o su grupo, que expresa el significado que la situación bajo evaluación tiene para ellos.

En este contexto, los resultados de la evaluación no son “hechos” en un sentido definitivo, sino más bien una “**creación**”, producto de una interacción que incluye tanto al evaluador como a otras personas involucradas en el proceso. Lo que surge de éste es una o más **construcciones** que constituyen la realidad del caso (Guba y Lincoln, 1992).

No obstante lo anterior, existe una tendencia generalizada, tanto en el plano nacional como en el internacional, a utilizar pruebas de rendimiento (escritas, orales o prácticas) para apreciar el nivel de conocimientos adquiridos por los estudiantes en el sistema escolar, lo cual indica que todavía

existe una fuerte corriente de personas que enfatizan los procedimientos objetivos en la medición de los aprendizajes. Así, por ejemplo, Ausubel, Novak y Hanesian (1983) han expresado que:

*Si nuestra actitud hacia la educación es realmente seria, tenemos que precisar la manera de medir los resultados del aprendizaje de los estudiantes como individuos y de averiguar si aquéllos concuerdan con nuestro propósitos educativos...Importa igualmente que [los procedimientos de evaluación] suministren datos que nos hagan posible vigilar la marcha del proceso educativo y asegurar así un control de calidad sobre la empresa en cuestión (p. 514).*

De lo anteriormente expuesto, se desprende que la evaluación del proceso educativo es una función compleja que comprende muchos aspectos (conocimientos, actitudes, aptitudes, habilidades, destrezas, creatividad, estrategias, valores, personalidad) y, por supuesto, no se agota en la utilización de las pruebas de rendimiento académico, como único procedimiento de apreciación de los resultados de los aprendizajes, sino que requiere de otros medios, tales como, por ejemplo, la observación directa de la conducta del alumno, escalas, inventarios, cuestionarios, entrevistas e informes personales, entre otros. En este capítulo se tratará lo referente a las pruebas de rendimiento académico y, particularmente, sobre las pruebas objetivas.

### **Definición y Propósitos**

Como ya se ha mencionado en la sección anterior de este mismo capítulo, la medición del aprendizaje, como parte de la función de evaluación, marcha correlativamente con el proceso de enseñanza-aprendizaje. En este sentido, hay tres momentos claves en los que se pone de manifiesto la importancia de

la medición-evaluación en el proceso instruccional. Ellos son: (a) antes de iniciar la enseñanza; (b) durante el proceso; y (c) al final del mismo.

Antes de iniciar el proceso de instrucción, el docente debe determinar las características cognitivas y afectivas de entrada de los estudiantes con el propósito de adecuar la enseñanza a sus necesidades específicas. Entre estas características, una de las más relevante es el nivel de conocimientos previos, como lo han expresado autores tales como Ausubel y cols (1983); y Klausmeier y Goodwin (1966), entre otros; por cuanto es necesario que la nueva información que adquirirá el alumno, se articule armónicamente con los conocimientos ya existentes en su estructura cognitiva, a objeto de lograr un aprendizaje significativo (evaluación diagnóstica). Este es uno de los casos en que se utilizan las pruebas de rendimiento académico (PRA).

Por otra parte, una vez iniciado el proceso de enseñanza-aprendizaje, el docente necesita disponer de ciertos mecanismos de control que les permitan saber en qué medida los alumnos progresan, a través de la secuencia de aprendizaje, a fin de introducir las medidas correctivas que fueren necesarias (evaluación formativa). Para este propósito, también se utilizan las pruebas de rendimiento académico.

Finalmente, una vez concluido, parcial o totalmente, el período académico, el docente debe determinar en qué medida los objetivos instruccionales propuestos fueron logrados (evaluación sumativa). En este caso, es igualmente apropiado utilizar las pruebas de rendimiento académico.

Pero, ¿qué son las pruebas de rendimiento académico? La respuesta a esta pregunta puede ser deducida de la exposición precedente. Al respecto, estas pruebas pueden ser definidas como procedimientos sistemáticos que utilizan los docentes con el fin de determinar el nivel de conocimientos de los

estudiantes en una disciplina determinada, antes, durante y al final de un período académico.

### **Características de las Pruebas de Rendimiento Académico**

Para que los resultados de las pruebas tengan algún sentido práctico, es necesario que cumplan con ciertos requisitos, como son los de validez, confiabilidad, discriminabilidad y factibilidad. Los aspectos conceptuales y procedimentales de los dos primeros requisitos (validez y confiabilidad) ya fueron expuestos en el Capítulo 2 de esta obra, por lo que no repetiremos estos conceptos en esta sección. No obstante, consideramos importante destacar el tipo específico de validez y confiabilidad que es significativo en el caso de las pruebas de rendimiento académico (PRA).

Para las PRA es particularmente relevante la validez de contenido, ya que es importante estar seguro de que los ítemes que integran la prueba realmente constituyen una muestra representativa del dominio de contenido (disciplina), que se desea medir; de lo contrario, mal se podría hacer una adecuada interpretación del desempeño de los alumnos con respecto al universo de contenido en cuestión.

Por ejemplo, si se desea elaborar una PRA para determinar el nivel de conocimientos que un grupo de alumnos tiene en aritmética (operaciones de suma, resta, multiplicación y división) y dicha prueba quedó constituida por 40 reactivos, tendrá validez de contenido en la medida que en dicho conjunto de ítemes se encuentren representadas las cuatro tipos de operaciones aritméticas y en cada caso los reactivos estén balanceados en cuanto a su nivel de dificultad.

Es decir, si se excluyera aunque fuera uno de los tipos de operaciones y, en alguno de los casos, los ítemes fueran más fáciles que en otros, la validez de contenido estaría seriamente cuestionada. De lo anterior, se deduce que el

aspecto clave de la validez de contenido es el muestreo. No olvidemos que una prueba siempre será una muestra de muchos reactivos (teóricamente infinitos) que podrían formularse en relación con un dominio de contenido determinado.

En definitiva, lo que interesa, desde el punto de vista de la validez de contenido, es la precisión con que la prueba mide el contenido de la materia y los objetivos instruccionales desarrollados durante el período académico. En tal sentido, para elaborar una PRA con una alta probabilidad de tener validez de contenido, se debería proceder de la siguiente manera: (a) identificar las áreas temáticas (contenidos) y los objetivos a ser evaluados; (b) elaborar una tabla de especificaciones, en la cual se identifique el número de ítems que son necesarios incluir en la prueba, por cada área temática y tipo de objetivo; (c) elaborar una prueba ajustada a lo previsto en la tabla de especificaciones; (d) someter la prueba a juicios de expertos (otros especialistas en la materia); y (e) como resultados de lo anterior, introducir los correctivos que fueran necesarios.

En relación con la confiabilidad, todo instrumento de medición debería mostrar evidencia, por lo menos de estabilidad en sus resultados; y del grado de homogeneidad de los reactivos. No obstante, para los casos de PRA no estandarizadas, el tipo de confiabilidad que más se reporta es el de consistencia interna, el cual permite estimar la homogeneidad de la prueba o el grado de intercorrelación de los ítems que conforman la misma. El procedimiento utilizado en estos casos es el propuesto por Kuder y Richardson (1937), ya descrito en el Capítulo 2, de esta obra.

En cuanto al requisito de discriminabilidad de una PRA, se trata de la capacidad del instrumento para distinguir al máximo a los sujetos que tienen un alto dominio de la materia, en relación con aquéllos cuyo dominio es bajo

o muy bajo. Esta característica depende en gran medida del nivel discriminativo de los ítemes componentes de la prueba y explica, a la vez que refleja, la confiabilidad y validez del instrumento.

Sin embargo, en cierto grado depende también de la distribución de puntajes totales y de si la prueba suministra un adecuado soporte superior para los sujetos más dotados del grupo. Por ejemplo, en el caso de la distribución normal, dado que la mayoría (68,26 por ciento) de las puntuaciones tienden a concentrarse en la mitad de la distribución, la mayor discriminación se logra en los extremos de la curva (puntajes altos versus puntajes bajos) donde hay menos calificaciones que en el centro.

Finalmente, la factibilidad, como requisito de las PRA, se refiere a la utilidad de los resultados para tomar decisiones importantes en relación con la retroalimentación diferencial de los estudiantes y profesores, así como sobre las fortalezas y debilidades relativas a la enseñanza y al aprendizaje. Por otra parte, una prueba factible debería estar adaptada en forma y contenido a la edad de los sujetos a quienes les será aplicada. Además, es necesario que la prueba sea económica en cuanto a costo y al tiempo necesario para administrarla, calificarla e interpretarla.

## Clasificación de las Pruebas de Rendimiento Académico

Existen diferentes tipos de pruebas para medir los conocimientos que los estudiantes obtienen en el sistema escolar. Estas pruebas han sido clasificadas por diferentes autores atendiendo a distintos criterios. En esta obra, se seguirá un punto de vista clásico, el cual se refiere a la manera como el estudiante responde. De acuerdo con este criterio, estos instrumentos pueden ser categorizados como pruebas escritas, orales y prácticas. Por otra parte, atendiendo a la formalidad en la construcción y a la orientación métrica, estos instrumentos pueden ser clasificadas en: (a) pruebas informales (construidas por el docente); (b) pruebas estandarizadas; (c) pruebas referidas a criterios. Esta clasificación será tratada en el capítulo 5 de esta obra.

### **Pruebas Escritas**

Las pruebas escritas, como su nombre lo indica, son instrumentos de lápiz y papel, en los cuales las preguntas formuladas por el docente son respondidas por los alumnos de algunas de las siguientes maneras: (a) identificando y marcando la respuesta; (b) construyendo la respuesta, la cual se expresa a través de un breve ensayo o composición; y (c) utilizando una combinación de las dos modalidades anteriores. Estas formas de responder la prueba escrita, permite clasificarlas en: objetivas, de ensayo y mixtas.

El hecho de ser escritas no garantiza, por sí mismo, que este tipo de pruebas sean mejores o peores que otras, en términos de su calidad y eficiencia. No obstante, presentan una ventaja importante con respecto a las orales, ya que las respuestas escritas permiten ser analizadas y calificadas de mejor manera que las orales. En cierto sentido, representan un testimonio de

lo que verdaderamente respondió el estudiante, con base en lo cual se puede justificar la calificación emitida, en caso de algún reclamo.

### **Pruebas Objetivas**

Las pruebas objetivas son aquellas en las que el estudiante no necesita construir o redactar la respuesta, sino leer la pregunta, pensar la respuesta, identificarla y marcarla; o leer la pregunta, pensar la respuesta y completarla. Son pruebas de respuestas breves; su mayor ventaja está en que se elimina la subjetividad y la variabilidad al calificarlas, ya que de antemano se establecen criterios precisos e invariables para puntuarlas; comúnmente se utiliza una clave de calificación que designa las respuestas correctas.

Por otra parte, los ítemes de las pruebas objetivas son seleccionados cuidadosa y sistemáticamente para que constituyan una muestra representativa del contenido abarcado y de las competencias evaluadas. Tal resultado se logra mediante la planificación detallada y consciente de la prueba, mediante la utilización de una tabla de especificaciones. En ésta se pueden relacionar directamente los contenidos y objetivos instruccionales y, en función de ello, decidir acerca de los ítemes que deben ser incluidos en la prueba.

Tomando en cuenta que no es posible comprobar la totalidad del conocimiento deseable dentro de un área dada, debe prestarse mucha atención a la selección de los ítemes, en el sentido de que éstos sean importantes (no triviales) y representativos de los hechos, conceptos, principios y aplicaciones particulares que se espera domine el estudiante.

Las pruebas objetivas están integradas por ítemes que se caracterizan por su brevedad y por la rapidez con que pueden ser respondidos. Este atributo constituye otra ventaja, por cuanto ello permite un muestreo más amplio y sistemático del conocimiento del que sería posible por otros medios. Algunos

autores consideran que aun cuando la capacidad de reconocer una alternativa de respuesta correcta no supone necesariamente la capacidad de recordarla espontáneamente, la correlación entre ambas es razonablemente buena (ver, por ejemplo, Plumlee, 1947; Tyler, 1934, citados por Ausubel y cols., 1983).

Una ventaja adicional de las pruebas objetivas consiste en que permite refinar los ítemes después de utilizarlos inicialmente, para aumentar la claridad y la discriminabilidad de los mismos y, con ello, la confiabilidad y validez de la prueba.

Las pruebas objetivas han recibido serias críticas, algunas de ellas justificadas, pero muchas otras basadas en la falta de comprensión de su naturaleza, funciones y limitaciones inherentes. Algunas de estas críticas son las siguientes:

1. Se argumenta y, a veces con razón, que las pruebas objetivas miden el reconocimiento mecánico de ítemes de conocimiento o de información referida a hechos, datos o fechas (primer nivel de la taxonomía de Bloom); que son relativamente triviales e inconexos, en lugar de la comprensión genuina de conceptos, principios y relaciones amplios, así como la capacidad de interpretar hechos y aplicar conocimientos.

2. Ciertamente, las pruebas objetivas tienen algunas limitaciones para medir conocimientos referidos a los procesos cognitivos de más alto nivel (v.g., evaluación). Sin embargo, cuando la persona que elabora la prueba sabe como hacerlo, en algunos casos se puede llegar a medir hasta capacidad de síntesis, además de las de análisis, aplicación y comprensión.

3. Algunas veces, debido a una mala elaboración de los ítemes, particularmente en lo que respecta a las alternativas de respuestas, la

opción correcta puede ser fácilmente identificada o deducida, sin que ello implique un verdadero conocimiento del asunto por parte del estudiante. Esta debilidad es subsanable si quienes desean usar este tipo de instrumento se preocupan en entrenarse para tal fin.

4. Existe una alta probabilidad de que algunos ítemes sean respondidos correctamente por simple azar. Esto es cierto, aun cuando también es verdad que existen procedimientos especiales para la corrección de este tipo de ítemes objetivos que toma en cuenta esta posibilidad.

5. Las pruebas objetivas tienden a favorecer la velocidad de algunos estudiantes frente a la capacidad de otros. lo ideal sería lo contrario, para lo se requeriría disponer de una gama amplia y cuidadosamente graduada con respecto a dificultad, con tiempo suficiente para que la mayoría de los estudiantes concluyera la prueba, y no incluyendo el doble de ítemes que el estudiante promedio tiene tiempo de contestar.

Aparte de las consideraciones anteriores, es importante estar consciente de que, ciertamente, las pruebas objetivas presentan limitaciones para medir ciertos tipos de aprendizajes. Por ejemplo, las pruebas de selección múltiple no pueden medir, por definición, la capacidad de los estudiantes para formular espontáneamente hipótesis pertinentes, para reconocer datos clínicos o de laboratorio que sean válidos, para reunir testimonios en favor de una afirmación, para planear un experimento original, para estructurar un argumento defendible o para realizar trabajos creativos; sin embargo existen otros tipos de instrumentos de medición para probar el logro de estos objetivos (Ausubel y cols., 1983). Las pruebas objetivas pueden estar integradas por ítemes de varios tipos; por ejemplo, verdadero-falso, pareo, completación, selección simple y múltiple.

**Ítemes verdadero-falso.** Estos ítemes se expresan en forma de sentencias o frases cortas que demandan una contestación dicotómica del tipo: verdadero-falso, si-no, acuerdo-desacuerdo, bien-mal y otras respuestas similares. Esta selección tiene especial importancia en actividades que permiten reducir la respuesta a dos alternativas o posibilidades significativas, por ejemplo, probar hipótesis, juzgar conclusiones, evaluar extrapolaciones y relaciones. Hurtado de Mendoza (1980) ha identificado 11 formas en que pueden presentarse las proposiciones o frases de una prueba del tipo verdadero-falso, las cuales se presentan a continuación.

Tipo de Item	V	F
<b>1. Hechos:</b> La distancia más cerca entre dos puntos es una recta.	x	<input type="checkbox"/>
<b>2. Generalización:</b> Todas las cuestiones que preguntan por ¿quién?, ¿qué?, ¿cuándo? o ¿dónde? pueden clasificarse apropiadamente como preguntas de información objetiva.	x	<input type="checkbox"/>
<b>3. Comparación:</b> Los pueblos andinos y costeños son similares en su temperamento. <i>Puede usarse también: tanto, ambos.</i>	<input type="checkbox"/>	x
<b>4. Condicional:</b> Si las medidas de tendencia central (media, mediana y modo) de una distribución son iguales, la distribución es normal. <i>También puede usarse: cuando, en vez de si...</i>	x	<input type="checkbox"/>
<b>5. Relación:</b> A mayor número de sujetos (seleccionados aleatoriamente) en una muestra, el error estándar de la media es mayor. <i>Puede usarse: a más alto... a más bajo incrementando... tiende a el peso de... depende de...</i>	<input type="checkbox"/>	x
<b>6. Explicativa:</b> La razón principal para utilizar pruebas objetivas en la valuación del rendimiento académico es su rapidez para ser contestada y calificada. <i>Puede usarse: el propósito de... es uno de los factores que... pesar de... desde... aunque... dado que...</i>	x	<input type="checkbox"/>
<b>7. Ejemplificación:</b> Un ejemplo de ciudad costera en Venezuela sería Coro.	x	<input type="checkbox"/>
<b>8. Evidencia:</b> Estudios recientes en neurociencia han determinado que la creatividad está determinada por los dos hemisferios cerebrales.	x	<input type="checkbox"/>
<b>9. Predicción:</b> Se puede esperar un incremento de 0,30 a 0,60 en el coeficiente de confiabilidad de una prueba si se duplica el número de ítems.	<input type="checkbox"/>	x
<b>10. Procedimiento:</b> Para calcular la media aritmética de una distribución de puntajes se deben sumar todos los valores de la distribución y dividirlos por el número total de valores. <i>Puede usarse también: con el fin de ... se debe un método... es un paso esencial... es el uso de... mediante... el primer paso...</i>	x	<input type="checkbox"/>
<b>11. Evaluación:</b> Una buena prueba de rendimiento académico incluirá más reactivos de hechos que generalizaciones. <i>También puede usarse: es mejor... que la mejor... es el método más fácil... es el mayor inconveniente... la máxima... es fácil... es difícil... es posible...</i>	x	<input type="checkbox"/>

Como se puede observar en los ejemplos anteriores, las pruebas integradas con ítemes del tipo falso-verdadero, pueden utilizarse para medir el conocimiento de relaciones funcionales, la comprensión de sucesos, ideas o principios y la habilidad para aplicarlos en la solución de problemas. Es más bien un planteamiento incorrecto del reactivo y no la forma del mismo, lo que puede dar lugar a que este tipo de prueba se limite a medir conocimientos memorísticos y datos aislados, como han criticado algunos autores.

La razón fundamental para utilizar pruebas con ítemes del tipo verdadero-falso es que proporcionan una manera simple y directa de medir el resultado esencial de la educación formal. El argumento a favor de la validez de los ítemes verdadero-falso en cuanto a medidas del rendimiento educacional ha sido resumido por Ebel (1977) en cuatro proposiciones, a saber:

1. Lo esencial del rendimiento educacional reside en el dominio de conocimientos verbales útiles.
2. Todo conocimiento verbal puede expresarse en proposiciones.
3. Una proposición es cualquier enunciado del que se puede decir que es verdadero o falso.
4. El grado de dominio que un estudiante posee sobre determinada área de conocimiento está indicado por su capacidad de juzgar la verdad o falsedad de las proposiciones relacionadas con él.

Tal vez una de las críticas más serias que se le hacen a este tipo de prueba es el riesgo que tienen de ser respondidas por azar. Sin embargo, en la práctica, tal limitación no parecería constituir un problema tan grave, si se toma en cuenta que la influencia de las contestaciones por azar tiende a

disminuir en la medida que aumenta la extensión de la prueba. Así tenemos que la probabilidad de responder correctamente por simple azar en una prueba de un ítem es del 50 por ciento; mientras que para dos ítemes la probabilidad es de 25 por ciento, en una prueba de cinco reactivos, aquella baja al 3 por ciento, y en una de diez, al 1 por ciento (ver Stanley, 1964).

A continuación se presentan algunas recomendaciones para la elaboración de los ítemes verdadero-falso:

1. Antes de escribir el ítem, conviene tener clasificado el objetivo en que se basa, de acuerdo con la clase de conducta que implica.

2. El ítem debe ser introducido con palabras que expresen la necesidad que tiene el alumno de seleccionar, organizar y aplicar conocimientos.

3. Los enunciados deben ser breves y de estructura sencilla.

4. En cada enunciado se debe incluir sólo una idea central significativa.

5. Los reactivos deben ser redactados de tal manera que se le pueda juzgar inequívocamente como verdadero o falso.

6. Se debe evitar la utilización de vocablos tales como: “siempre”, “nunca”, “todo”, “ninguno” y “solamente”; éstos tienden a ser usados en ítemes falsos. Asimismo, se debe evitar el uso de términos como: “usualmente”, “acaso” y “a veces”, los cuales tienden a ser utilizados en ítemes verdaderos.

7. Se debe usar cuidadosamente los enunciados negativos y se debe evitar la doble negación.

8. Se debe incluir en la prueba igual número de ítemes verdaderos que de reactivos falsos.

***Ítemes de pareo.*** Llamados también reactivos de correspondencia o de

asociación. En este tipo de prueba, se le presentan al estudiante dos tipos de listas o columnas (A y B) de nombres, hechos o principios. En una de las columnas (A) se escribe la lista de **premisas**; mientras que en la otra (B) se escriben las alternativas de respuestas. A continuación se presenta un ejemplo de este tipo de ítem, adaptado de Gronlund (1976).

*Ejemplo 4*

<b>Instrucciones:</b> La columna A es una lista de características de los ítemes objetivos. En los paréntesis que aparecen a la izquierda de cada enunciado escriba la letra del ítem de la columna B que se ajuste mejor al enunciado. Cada respuesta de la columna B se puede usar una sola vez, más de una vez, o no usarse.		
	<b>Columna A</b>	<b>Columna B</b>
(D)	1. El mejor para medir habilidades de calcular.	A. Ítem de pareo
(C)	2. El menos útil para el diagnóstico educativo.	B. Ítem de opción múltiple
(B)	3. Mide la mayor variedad de resultados de aprendizaje.	C. Ítem verdadero-falso
(D)	4. El más difícil de calificar objetivamente.	D. Ítem de completación
(C)	5. Da lugar a la mayor puntuación lograda sólo por azar.	
(D)	6. Mide el aprendizaje a nivel de memorización	

Como se puede observar en el ejemplo anterior, la alternativa A de la columna B no fue utilizada en ningún caso, la opción B usada una vez, la alternativa C, dos veces y la opción D, tres veces. Sin embargo, existen otras

situaciones en las cuales cada alternativa de respuesta se usa una sola vez. En estos casos, el número de opciones de la columna B (derecha) debe ser mayor que las de la columna A (izquierda). A continuación se ilustra este planteamiento con un ejemplo.

### Ejemplo 5

**Instrucciones:** La columna A contiene una lista de estados de la República de Venezuela. Escriba en los paréntesis que aparecen del lado izquierdo de cada enunciado, la letra del ítem de la columna B que se ajuste mejor al enunciado. Cada respuesta de la columna B puede ser usada sólo una vez.

	<b>Columna A</b>	<b>Columna B</b>
E)	1. Miranda	A. Distrito Federal
D)	2. Lara	B. Barcelona
F)	3. Falcón	C. San Cristobal
C)	4. Táchira	D. Barquisimeto
B)	5. Anzoategui	E. Los Teques
		F. Coro
		G. San Carlos

Algunos autores consideran que los ítemes de pareo son muy útiles en la evaluación del rendimiento académico, ya que a través de ellos se pueden presentar una gran cantidad de contenido (ver Nunnally, 1972; Gronlund, 1976; Ebel, 1977) No obstante, para su uso apropiado se deben seguir las siguientes recomendaciones:

1. Las instrucciones deben ser claras y deben indicar cómo se hará el pareo de los ítemes de las columnas.
2. Se debe incluir material relevante y homogéneo en cada ítem.

3. No se debe incluir más de seis premisas en la columna A.

4. La columna B debe contener dos o tres alternativas de respuestas más (distractores) de las que realmente se necesitan para contestar todas las premisas.

**Ítemes de completación.** Se identifican también como ítemes de respuestas breves. En este tipo de reactivo el estudiante no elige la respuesta correcta dentro de un conjunto de opciones, sino que debe proporcionarla o escribirla en un espacio en blanco previsto para tal fin. El ítem consta de una pregunta o enunciado incompleto al que el estudiante responde proporcionando las palabras, número o símbolos adecuados. A continuación se presenta un ejemplo que ilustra este tipo de ítem.

#### *Ejemplo 6*

1. De acuerdo con lo establecido en la Constitución de la República de Venezuela, para ser presidente se requiere tener como mínimo \_\_\_\_\_ años de edad.

Este tipo de ítem ocupa un puesto intermedio entre los ítemes de objetivos y los de ensayo, ya que en lugar de seleccionar la respuesta correcta el sujeto debe construir la respuesta. Sin embargo, se le considera como un ítem objetivo porque uno o más términos deben ser proporcionados, los cuales pueden ser anticipados previamente para fines de calificación.

La utilización de ítemes de respuestas breves presenta dos dificultades principales; primero, resulta bastante laborioso redactar el enunciado incompleto de manera que sólo haya una única respuesta correcta, ya que los estudiantes pueden utilizar sinónimos, u otros términos que tengan significados relacionados con la idea expresa en el ítem. El segundo se refiere

al problema de la ortografía; es decir, cuando la respuesta proporcionada es correcta pero está mal escrita. En este caso hay que decidir si se le da o no crédito a la respuesta.

De acuerdo con Gronlund (1976) este tipo de ítem se debe utilizar sólo en situaciones especiales en las que el suministro de la respuesta sea parte necesaria del resultado de aprendizaje que se va a medir; por ejemplo, cuando el objetivo es que los estudiantes recuerden la información, cuando se utilizan problemas de cálculo numérico o cuando un ítem de selección haría que la respuesta fuese obvia.

Algunas de las recomendaciones para la redacción de ítemes de respuestas breves son las siguientes:

1. No se debe usar más de uno o dos espacios en blanco.
2. Se debe prever que haya un solo término que satisfaga la respuesta para cada espacio en blanco a ser completado.
3. Cuando se redacte el ítem se debe prever que la respuesta a ser proporcionada corresponda a términos relevantes en relación con el contenido que se pretende evaluar.
4. En la medida de lo posible, los espacios en blanco deben estar ubicados hacia el final de la frase.
5. Se debe evitar hacer transcripciones textuales del material en la construcción del ítem.
6. Se deben evitar claves ajenas a la respuesta.

***Ítemes de selección simple y múltiple.*** Este tipo de ítem consta de dos partes: (a) un enunciado, que puede estar representado por una frase o una pregunta; y (b) cuatro o más opciones de respuestas, una de las cuales es la opción correcta (en el caso de los ítemes de opción simple), el resto se

conocen como **distractores** y deben guardar relación con el enunciado que las introduce.

Cuando la selección supone escoger más de una alternativa de las varias que se presentan, se dice que item es de opción múltiple y, en este caso se incluye en su redacción un número mayor de distractores. A continuación se presentan dos ejemplos que ilustran los ítemes de opción simple y los de selección múltiple.

*Ejemplo 7. Item de opción simple*

**Instrucción:** Marque con una equis (x) en los paréntesis de la izquierda la alternativa que usted considere es la respuesta correcta.

En una situación en que el elemento A es mayor que B y B es mayor que C, la relación entre A y C es:

1.  A es igual que C
2.  A es menor que C
3.  A es mayor que C
4.  No se puede saber

*Ejemplo 8. Item de opción múltiple*

**Instrucción:** Marque con una equis (x) las alternativas que conforman la respuesta correcta.

Las partes en que se divide el cuerpo humano son:

1	<input type="checkbox"/>	Ojos	3	<input checked="" type="checkbox"/>	Tronco	6	<input type="checkbox"/>	Boca
2	<input checked="" type="checkbox"/>	Cabeza	4	<input type="checkbox"/>	Nariz	7	<input type="checkbox"/>	Manos
3	<input type="checkbox"/>	Brazos	5	<input type="checkbox"/>	Pies	8	<input checked="" type="checkbox"/>	Extremidades

Los ítemes de opción simple y múltiple son los más utilizados por los

especialistas en construcción de pruebas, a pesar de ser el tipo de reactivo que es más exigente en cuanto a su redacción. Esta preferencia se explica por varias razones:

1. Es uno de los tipos de ítemes objetivos más confiables; ya que permite controlar al máximo la respuesta por azar.

2. Permiten medir los aspectos relevantes de la mayoría de los objetivos instuccionales, por ejemplo, conocimientos de hechos, comprensión de lectura, aplicación de principios, relación de causalidad, prueba de hipótesis, analogías, categorización, análisis y síntesis.

3. Además, ayuda a distinguir a los estudiantes que tienen alto rendimiento de aquellos cuyo desempeño es bajo.

4. En general, este tipo de reactivo, se presta menos a la ambigüedad y a las falsas interpretaciones.

A continuación se presentan algunas recomendaciones para redacción de los ítemes de opción simple y múltiple, éstas son:

1. El ítem debe referirse a un aspecto importante o relevante del contenido que se desea evaluar.

2. Los distractores deben plantearse en el contexto del área de contenido que se está evaluando.

3. Las alternativas de respuestas correctas no deberían ser demasiado diferentes en apariencia en comparación con los distractores.

4. Las alternativas de respuestas deberían ser ordenadas al azar en cada ítem.

5. Se deben evitar alternativas tales como: “ninguna de las anteriores”, “Ambas a y c” o “todas las anteriores”.

6. Cada ítem debe ser independiente de los demás.

7. Se debe presentar sólo un problema en el enunciado o pie del

item.

8. El ítem se debe redactar en un lenguaje claro y sencillo.

9. El enunciado del ítem debe ser redactado en forma afirmativa, siempre que sea posible.

10. Hay que asegurarse de que la respuesta que se pretende es la correcta o, claramente, la mejor.

11. Se debe evitar el uso de claves verbales que permitan a los estudiantes seleccionar las respuestas correctas o eliminar una opción incorrecta.

12. Se debe tratar que los distractores aparezcan como posibles y atractivos para el estudiante poco informado.

13. Se debe controlar la dificultad del ítem ya sea variando el problema en el pie del ítem o cambiando las opciones.

14. Se debe hacer variar la extensión relativa de la respuesta correcta para eliminar la longitud como una posible clave.

***Normas para puntuar las pruebas objetivas.*** Una de las aparentes desventajas que presentan las pruebas objetivas, diseñadas para evaluar el rendimiento escolar, es que sus resultados tienden a estar influenciados por el tipo de respuesta por azar o adivinación. De allí que los especialistas en el campo de evaluación de los aprendizajes recomendaran, en un primer momento, la utilización de algunos procedimientos que permitieran controlar el efecto de la respuesta aleatoria en este tipo de prueba (ver, por ejemplo, Anastasi, 1976; Sach, 1970; Nunnally, 1972).

Para corregir el efecto de la adivinación de las respuestas en las pruebas objetivas, es necesario sustraer del puntaje de la prueba una cantidad igual a la ganancia que se supone puede ser lograda por simple azar al responder. En el caso de las pruebas diseñadas con ítems del tipo verdadero-falso, se puede

esperar que por cada respuesta incorrecta el sujeto emita una correcta. De allí que para su calificación se sustraiga las respuestas incorrectas de las correctas, por ejemplo, en el caso de un estudiante que respondió 10 preguntas correctas y 10 incorrectas en una prueba de este tipo, tendrá cero (0) como puntaje final.

Si las preguntas presentan tres respuestas igualmente posibles en vez de dos, se debe esperar que el estudiante de dos respuestas incorrectas por cada respuesta correcta. En tal caso habrá de sustraerse la mitad del número de respuestas erróneas del número total de respuestas acertadas para neutralizar la adivinación. Si los ítemes de respuestas múltiples incluyen cinco opciones de respuestas para cada pregunta, una sola de las cuales es correcta, la proporción esperada de respuestas correctas será de 4 por 1, y la corrección deberá hacerse sustrayendo un cuarto del número de respuestas erróneas del número de respuestas correctas. Este razonamiento ha llevado a la utilización de una fórmula general para todos estos casos, la cual se presenta más abajo:

$$Pc = c - \frac{I}{N-1}$$

Donde:

Pc es el puntaje corregido

C es el número de respuestas correctas

I es el número de respuestas incorrectas

N es número de alternativas de respuesta en el ítem

Así, un estudiante que obtenga 15 respuestas correctas y 5 incorrectas en una prueba de 20 preguntas, con cuatro alternativas de respuesta, tendrá 13 como puntuación.

Otra manera de resolver el problema anterior, consiste no en castigar al estudiante que adivina, sino en neutralizar el efecto del azar recompensando

al que se abstiene de adivinar. Es decir, en vez de sustraer 50 unidades del puntaje del estudiante que intenta adivinar, se podría sumar 50 unidades al que no recurre a la adivinación. De esta manera también se elimina la ventaja que puede esperarse de la respuesta por azar.

En el caso anterior, se parte del supuesto de que si el estudiante que no adivinó lo hubiera hecho, habría dado una respuesta correcta a una mitad de los ítemes verdadero-falso. En ítemes de tres opciones, habría acertado una tercera parte de los reactivos. Este razonamiento ha llevado a la utilización de una segunda fórmula general para corregir los puntajes de las pruebas objetivas, cuando los sujetos se abstienen de responder aquellas preguntas que no saben o no están completamente seguros de sus respuestas, como se indica más abajo:

$$P_c = c + \frac{O}{N}$$

Donde:

$P_c$  es el puntaje corregido

$C$  es el número de respuestas correctas

$O$  es el número de preguntas omitidas

$N$  es el número de opciones por ítem

De esta manera, un estudiante que obtuvo 15 respuestas, dos incorrectas y tres omitidas, en una prueba de 20 reactivos de cuatro alternativas de respuestas por ítem, tendría 16 como calificación. Los resultados de investigaciones posteriores han sugerido que no existe una diferencia marcada entre los puntajes corregidos y los no corregidos, en cuanto a la posición relativa que ocupan los estudiantes de acuerdo con su desempeño académico. En este sentido, Ebel (1977) ha reportado que los estudios experimentales realizados sobre el efecto que el anuncio de la corrección de la

adivinación antes de tomar las pruebas tiene sobre la confiabilidad y validez de los puntajes de éstas han demostrado que es escaso o ninguno el mejoramiento atribuible a la corrección.

Ebel observó que la correlación entre los puntajes corregidos y no corregidos de las mismas respuestas, normalmente es igual o mayor que 0,90. De allí que actualmente se considere a las fórmulas de corrección por azar como algo poco relevante a los fines de perfeccionar la calificación de los estudiantes en las pruebas objetivas.

***Ponderaciones diferenciales de los reactivos.*** Normalmente, las puntuaciones de los alumnos en las pruebas de aprovechamiento se obtienen sumando las calificaciones individuales de los ítemes. El sistema más usado de ponderación consiste en otorgar un punto por cada respuesta correcta y cero punto por los reactivos respondidos incorrectamente.

No obstante lo anterior, algunos especialistas en construcción de pruebas consideran que hay ciertas preguntas que deberían tener una ponderación mayor, por considerar que son más difíciles, más importantes, de mejor calidad técnica, más complejas o que requieren de más tiempo para ser respondidas. Por ejemplo, algunos investigadores consideran que las pruebas integradas por ítemes del tipo verdadero-falso deberían ser ponderados con un punto por cada respuesta correcta; mientras que otros, como los de selección múltiple, deberían ser ponderados con dos o más puntos.

El razonamiento anterior aunque pareciera lógico no ha demostrado, en la práctica, que tal ponderación diferencial de los ítemes tenga algún efecto positivo en el mejoramiento de la confiabilidad y validez de las pruebas de rendimiento académico, aunque tampoco la hacen peor (Ebel, 1977).

### **Pruebas de Ensayo**

En los casos en que se requiere la evaluación de aprendizajes complejos y se

desea apreciar la capacidad del sujeto para producir, organizar y expresar sus ideas en forma escrita; y habilidad de razonamiento independiente, se prefieren los reactivos de ensayo en vez de los ítemes objetivos. Los ítemes de ensayo, también llamados de composición, son aquellos en que, formulada la pregunta, el sujeto tiene completa libertad para responder; es decir, el individuo queda en condición de decidir cómo enfocar el problema, qué información utilizar, cómo organizar la respuesta y qué grado de importancia dar a cada aspecto de la respuesta. Algunos autores recomiendan utilizar las pruebas de ensayo cuando:

1. El grupo es pequeño y la prueba no se volverá a utilizar con los mismos alumnos;
2. El instructor desea mejorar y alentar la habilidad del estudiante para expresarse por escrito;
3. Se está interesado tanto en explorar actitudes como en medir conocimientos;
4. El profesor confía más en su propia capacidad de lectura crítica que en la de diseñador imaginativo de preguntas; y
5. Cuando se cuenta con poco tiempo para preparar la prueba (ver Lindvall, 1967; Chase, 1968).

Las diferencias básicas que existen entre las pruebas tipo ensayo y las objetivas, de acuerdo con Hurtado de Mendoza (1980), se presentan en el Cuadro 24.

*Cuadro 24. Diferencias entre las pruebas tipo ensayo y las objetivas*

Prueba Tipo Ensayo	Prueba Objetiva
<ul style="list-style-type: none"> <li>• Requiere que el estudiante planee su respuesta con sus propias palabras.</li> </ul>	<ul style="list-style-type: none"> <li>• Requiere que el estudiante escoja la respuesta entre dos o más alternativas.</li> </ul>
<ul style="list-style-type: none"> <li>• Contiene pocas preguntas, bastante generales que exigen contestación extensa.</li> </ul>	<ul style="list-style-type: none"> <li>• Contiene una amplia variedad de preguntas específicas que exigen contestación breve.</li> </ul>
<ul style="list-style-type: none"> <li>• Destaca y delimita sólo algunos aspectos importantes de un tema.</li> </ul>	<ul style="list-style-type: none"> <li>• Permite cubrir una gama más amplia de los aspectos de la materia a medir.</li> </ul>
<ul style="list-style-type: none"> <li>• Los estudiantes tienen que pensar y escribir.</li> </ul>	<ul style="list-style-type: none"> <li>• Los alumnos usan la mayor parte del tiempo en leer y pensar.</li> </ul>
<ul style="list-style-type: none"> <li>• La calidad depende de la capacidad para leer del examinador.</li> </ul>	<ul style="list-style-type: none"> <li>• La calidad depende de la habilidad de quien diseña la prueba.</li> </ul>
<ul style="list-style-type: none"> <li>• Se prepara fácilmente, pero es difícil de calificar y lleva tiempo.</li> </ul>	<ul style="list-style-type: none"> <li>• Su preparación es difícil y lleva a tiempo, pero se corrige con facilidad.</li> </ul>
<ul style="list-style-type: none"> <li>• Asegura mayor libertad al estudiante para expresar sus conocimientos e individualidad.</li> </ul>	<ul style="list-style-type: none"> <li>• Asegura mayor libertad a quien la construya para expresar sus conocimientos y valores.</li> </ul>
<ul style="list-style-type: none"> <li>• Las tareas del estudiante y las bases de medición aparecen menos clara.</li> </ul>	<ul style="list-style-type: none"> <li>• Establece más claramente las tareas a realizar y las bases de la medición.</li> </ul>
<ul style="list-style-type: none"> <li>• Permite y, en ocasiones, da pie a la improvisación o mera palabrería.</li> </ul>	<ul style="list-style-type: none"> <li>• Permite y, en ocasiones da pie a que se adivine.</li> </ul>
<ul style="list-style-type: none"> <li>• La distribución de las medidas depende, en gran parte, del examinador.</li> </ul>	<ul style="list-style-type: none"> <li>• La distribución de las medidas la determina la prueba.</li> </ul>

***Tipos de ítems de ensayo.*** Existen básicamente dos tipos de reactivos de ensayo o composición: los de respuesta restringida y los de respuesta libre. En el primer caso se le pide al estudiante que de una respuesta breve y definida con precisión.

Por lo general, la pregunta incluye expresiones tales como:

“Enumere...”, “Defina...”, “De razones...”, “Describa...”; A continuación se presentan algunos ejemplos ilustrativos.

### *Ejemplo 9*

Enumere cinco diferencias entre las pruebas tipo ensayo y las objetivas.

### *Ejemplo 10*

Describa las ventajas que tienen las pruebas objetivas para evaluar el rendimiento académico, en comparación con otro tipo de pruebas. Limite su respuesta a una página.

### *Ejemplo 11*

El profesor Pérez, de la cátedra de evaluación, desea medir en sus estudiantes la “habilidad para diseñar y validar una prueba de rendimiento”.

- 1.** Describa los pasos que debe seguir el profesor Pérez.
- 2.** Indique las razones que justifican cada paso.

El profesor Pérez, de la cátedra de evaluación, desea medir en sus estudiantes la “habilidad para diseñar y validar una prueba de rendimiento”.

1. Describa los pasos que debe seguir el profesor Pérez.
2. Indique las razones que justifican cada paso.

El uso de los ítemes de respuestas restringidas tienen la ventaja que se pueden preparar más fácilmente, relacionarse más directamente con los resultados específicos del aprendizaje y calificarse cómodamente. Pero su principal desventaja es que no dan mucha oportunidad para organizar,

integrar y desarrollar patrones de respuesta esencialmente nuevos. En consecuencia, este tipo de reactivos son útiles para medir los resultados del aprendizaje en los niveles de comprensión, aplicación y análisis (taxonomía de Bloom), pero tienen relativamente poco valor para medir los niveles de síntesis y evaluación.

En el segundo caso, se le concede al sujeto amplia libertad en cuanto a la naturaleza y el alcance de su respuesta, aunque en algunos casos se pueden imponer límites mínimos de tipo prácticos, como los temporales, o de número de páginas. En consecuencia, el estudiante tiene suficiente libertad para demostrar sus capacidades de síntesis y de evaluación, imponiéndole un control necesario para asegurar que la pregunta pondrá de manifiesto las habilidades cognitivas deseadas. A continuación se presentan algunos ejemplos ilustrativos.

### *Ejemplo 12*

Elabore un plan para evaluar el rendimiento académico del curso que usted enseña. Asegúrese de incluir los procedimientos a utilizar, los materiales necesarios y justifique las razones que usted tuvo para hacer su selección.

### *Ejemplo 13*

*Asuma que se le ha proporcionado al estudiante una prueba completa de rendimiento académico, con errores en las instrucciones, en los ítems y en la disposición de éstos.*

Usando como criterio de evaluación las reglas y normas, para la elaboración de test, descrita en su libro de texto, emita un juicio crítico de la prueba. Incluya un análisis detallado de las fortalezas y debilidades de la misma, así como una evaluación sumaria de su calidad total y de su probable ficacia.

*Adaptado de Gronlund, 1976*

***Recomendaciones para su elaboración.*** Escribir buenos ítemes de ensayo no es una tarea fácil; requiere conocimiento y práctica. Los especialistas han recomendado algunas reglas que, aun cuando no resuelven totalmente el problema, ayudan en el propósito de escribir reactivos de composición sin ambigüedad y que pongan de manifiesto los conocimientos que se desean evaluar en el alumno (Nunnally, 1972; Gronlund, 1976; Hurtado de Mendoza, 1980). Algunas de estas reglas son:

1. Se debe utilizar un número relativamente grande de ítemes de respuestas restringidas, en vez de pocos reactivos de respuestas libres. Ello permite un mejor muestreo de contenido y, al mismo tiempo, facilita la calificación de la prueba debido a la semi-estructura que lleva implícita la prueba de respuestas restringidas.

2. Es necesario proporcionar suficiente información en la formulación de la pregunta, a fin de orientar al estudiante hacia la respuesta esperada. Algunas veces este tipo de preguntas son tan generales, que dificultan al sujeto precisar que se espera de ellos en las respuestas, lo cual podría inducir a contestaciones bajas o imprecisas.

3. Todos los estudiantes deben responder las mismas preguntas, si se quiere una medición efectiva. Algunos docentes acostubran a formular una lista de preguntas abiertas o de respuestas libres, para luego pedir a los estudiantes que contesten un subconjunto de ellas; por ejemplo, cinco ítemes para responder tres. En tal situación, existen nueve posibles combinaciones que los sujetos pueden hacer para responder, lo cual con toda seguridad dificultará la comparación de las respuestas, lo cual, a su vez, influirá en la precisión y objetividad en la calificación de los reactivos.

4. No deben formularse preguntas que sólo requieran de una

expresión de opinión del sujeto que responde. Las respuestas basadas en las opiniones de los respondientes no permiten tener un criterio común de calificación de las preguntas, ya que cada opinión no puede ser evaluada sino en función de las argumentaciones lógicas de cada individuo, las cuales pueden ser tan variadas como personas hayan.

5. Los reactivos deben ser fraseados de tal manera que estimulen al estudiante para responder con razonamiento de alto nivel cognitivo. Al respecto, en la formulación de las preguntas se deben emplear términos o expresiones, tales como, por ejemplo, “comparar”, “relacionar”, “explicar”, “analizar”, “evaluar”, “inferir”, “realizar un análisis comparativo de...”, “describir el proceso de...”, “realizar un análisis crítico de...”.

6. Se debe evitar comenzar la redacción de los reactivos de ensayo con palabras tales como: “qué”, “quién”, “cómo”, “dónde”, “mencione”, “enumere”, entre otros. Estos términos tienden a limitar la respuesta a resultados de conocimiento (primer nivel de la taxonomía de Bloom (Bloom, Hastings y Madaus, 1971).

***Reglas para calificar las pruebas de ensayo.*** Una de las limitaciones más importantes de las pruebas de ensayo está relacionada con la subjetividad del docente-evaluador en el momento de calificar las respuestas. La subjetividad, en este contexto, está relacionada con los sentimientos del calificador, los cuales pueden interferir con los juicios que hace sobre la calidad de las respuestas.

Tales sentimientos pueden expresarse en sesgo o prejuicio del evaluador con respecto a la persona que escribió el ensayo, con respecto al estilo del que escribe, el enfoque teórico utilizado u otros factores como la ortografía o la ininteligibilidad de la prueba. A continuación se presentan algunas de las

recomendaciones que han formulado diferentes especialistas, a fin de controlar al máximo las debilidades de las pruebas tipo ensayo.

1. Las respuestas de tipo restringido deben ser evaluadas por el método de puntos, usando una respuesta modelo como guía. Es decir, se consideran las diferentes partes de la pregunta y cada una se le asigna un puntaje.

2. Las respuestas libres deben ser calificadas mediante el método de clasificación numérica, usando como guías criterios claramente definidos. En este caso, se leen las pruebas y, de acuerdo con el nivel de calidad, se agrupan en cinco pilas con calificaciones que van del 1 al 5, o con letras comprendidas entre A y E, ambas inclusive.

3. Se deben evaluar todas las respuestas dadas por el grupo a una pregunta, antes de pasar a otro reactivo, en lugar de hacer la evaluación estudiante por estudiante. Ello permite mantener una norma más uniforme para juzgar las respuestas, y ayuda a controlar el “efecto halo” en la calificación.

4. Se debe evaluar anónimamente las respuestas que los estudiantes dan a las preguntas de ensayo. Es importante que el conocimiento (positivo o negativo) que el docente-evaluador tiene del estudiante no influya en la calificación de las respuestas. Para controlar este factor personal se recomienda evaluar los reactivos sin identificar previamente la prueba.

5. Se recomienda evaluar las respuestas a las preguntas de ensayo en función del aprendizaje que se está midiendo. Al calificar las respuestas se debe evitar, en la medida de lo posible, tomar en cuenta información factual que, aunque interesante, no venga al caso; así como, el estilo en que se escriba el ensayo y los errores de ortografía y gramaticales.

6. Se recomienda que, de ser posible, cada respuesta sea calificada por dos o más personas. Esto permitiría tener bajo control la confiabilidad de la prueba, al poderse comparar los resultados de dos o más jueces independientes.

Limitaciones de las pruebas de ensayo. Algunas de las críticas que comúnmente se le hacen a este tipo de pruebas son las siguientes:

1. Presenta un muestreo bastante limitado del área de contenido a ser evaluado.

2. Otro defecto se refiere a la naturaleza de la respuesta. Como el sujeto debe escribir las respuestas con sus propias palabras, su capacidad para escribir tiende a alterar la calificación.

3. La alta probabilidad de que la calificación a la prueba sea asignada de manera subjetiva, a pesar de las previsiones que se tomen al respecto para controlar tal aspecto. Bajo estas condiciones, asignar las puntuaciones a las preguntas se transforma en un trabajo laborioso, que consume bastante tiempo y las puntuaciones de los ítems tiende a ser poco estable de un evaluador a otro o aun en un mismo evaluador, cuando revisa la prueba en diferentes ocasiones.

En relación con lo anterior, Gronlund (1976) ha señalado que las limitaciones de las pruebas de ensayo son tan grandes que probablemente ningún maestro debiera usarla como la única medida de rendimiento académico. Pero, por otra parte, él cree que, este tipo de ítem es tan importante para medir ciertos resultados de aprendizaje que tal vez la mayoría de los maestros deberían usarlas en cierto grado.

### **Pruebas Orales**

Las pruebas orales son aquellas mediante las cuales los estudiantes responden de viva voz a las preguntas formuladas, de la misma manera, por los

examinadores (Lemus, 1974). Este tipo de pruebas, de gran demanda en el pasado, tienden a usarse cada vez menos como instrumentos confiables de evaluación escolar, aun cuando no podría decirse que han desaparecido del todo. Las pruebas orales se utilizan actualmente en situaciones específicas en las que interesa explorar los conocimientos en áreas tales como lenguaje, ciencias sociales y personalidad, entre otros.

En general, a través de estas pruebas se pueden apreciar aspectos personales que difícilmente se pueden lograr a través de otros medios, como apariencia personal, modales, desenvolvimiento, capacidad de argumentación, facilidad de expresión, impresión causada en otras personas.

Además, si se aplica para comprobar el dominio de conocimientos o habilidades intelectuales, permite la flexibilidad de la que suelen carecer los exámenes escritos: se puede, por ejemplo, pedir al examinando que amplíe, aclare, justifique razone o profundice su respuesta.

***Normas para su aplicación.*** Con el propósito de controlar, en la medida de lo posible, algunas de las debilidades que presentan las pruebas orales como instrumentos de evaluación, los especialistas en el área han formulado algunas recomendaciones, tales como las siguientes:

1. El uso de la prueba debe estar debidamente justificado, en función del objetivo que se pretenda evaluar de la naturaleza del contenido curricular.

2. Es necesario que el examen sea estructurado; es decir, que las preguntas estén previamente establecidas y que a los estudiantes se le formulen las mismas preguntas.

3. Los criterios de evaluación de las preguntas deben ser, igualmente, previamente establecidos, a fin de controlar al máximo la subjetividad del examinador.

4. Los resultados de la evaluación deben registrarse tan pronto como sea posible, a fin de evitar pérdida de información debido al fenómeno del olvido.

5. De ser posible, el juicio evaluativo sobre las repuestas debería ser el resultado de la decisión de un jurado (3 ó 4 jueces o especialistas) y no de la opinión de una sola persona.

***Limitaciones de las pruebas orales.*** Este tipo de pruebas, si se usan apropiadamente y en el contexto de ciertos tipos de aprendizajes, en el cual su empleo está plenamente justificado, pueden juzgar un buen papel como instrumentos de evaluación. Sin embargo, presentan algunas limitaciones, las cuales deben ser tomadas en cuenta en el caso de que se decida su utilización. Algunas de estas limitaciones se indican a continuación:

1. El empleo de las pruebas orales dificulta la obtención de resultados consistentes en un tiempo razonable.

2. Su carácter individual, exige el empleo de bastante tiempo como cada sujeto (30 minutos o más), si se quiere tener una muestra válida de sus conocimientos;

3. El contacto personal, que es una de sus ventajas, puede dar lugar al prejuicio y la parcialidad por parte del examinador y a que la tensión altere el proceso de razonamiento de quien presenta el examen;

4. No se ha demostrado que una prueba oral pueda medir mejor rasgos intangibles o mal definidos, tales como, carácter, capacidad, fuerza y otros.

### **Pruebas Prácticas**

Una manera de clasificar los tipos de conocimientos que se imparten en el sistema escolar es agrupándolo en: (a) declarativo, el cual se centra en el conocimiento acerca del qué; es decir, se refiere a la información específica

que posee el sujeto en un área determinada, lo cual incluye aspectos tales como: vocabulario, conceptos, principios, reglas y aspectos teóricos en general; (b) procedimental, tiene que ver con el conocimiento acerca del cómo hacer, a la acción práctica, al desempeño real y concreto; y (c) contextual, el cual se refiere al cuándo hacer qué; es decir, expresa el conocimiento discriminativo acerca la oportunidad en que se debe aplicar un determinado conocimiento o estrategia y no otra, dependiendo de las circunstancias.

Las pruebas prácticas son las indicadas para evaluar el nivel de conocimiento procedimental; o sea, la competencia adquirida para desempeñarse en una determinada asignatura o área del conocimiento. Este tipo de instrumentos reciben también el nombre de pruebas de ejecución o funcionales. Por lo general, se emplean para medir habilidades determinadas tales como la destreza manual, la habilidad mecánica, la habilidad artística y otras.

En las pruebas prácticas, se le solicita al sujeto realizar una determinada actividad o tarea, a través de la cual se intenta determinar su competencia o maestría para llevarla a cabo. Estas prácticas se aplican fundamentalmente en áreas tales como: ciencias (trabajos de laboratorio, incluyendo el uso de aparatos, herramientas, instrumentos y materiales); artes (canto, ejecución de instrumentos musicales, pintura, modelado); educación física (calistenia, deportes); artes industriales (carpintería, electricidad, mecánica, etc.); economía doméstica (cocina, decoración, cuidado de la ropa, costura); dactilografía, mecanografía y otras (Lafourcade, 1969).

La comprobación del nivel de competencia alcanzada para ejecutar una determinada tarea, toma en cuenta tanto el proceso (secuencia de pasos necesarios para realizarla) como el producto que se obtiene de aquél, aunque

no siempre es posible hacer esta separación en la práctica. Tales resultados pueden ser de presencia estable en el tiempo, o de apreciación momentánea. En el primer caso, los resultados pueden ser analizados las veces que sean necesarias sin que cambien en su apariencia física; por ejemplo: una reacción obtenida en un laboratorio, un vestido en economía doméstica, un mueble hecho en un taller de carpintería y otros.

En el segundo caso, los procesos y productos se dan simultáneamente y son de apreciación directa e inmediata, por ejemplo: una pieza musical interpretada con el piano; la interpretación vocal de una canción, el desempeño de un estudiante de medicina en una intervención quirúrgica, el desempeño en una prueba deportiva, bailar, nadar, jugar fútbol entre otras (Moredock, 1957, citado por Lafourcade, 1969).

Ryans (1951) considera que existen, por lo menos tres tipos de pruebas prácticas, las cuales se describen a continuación:

1. Aquellas en las cuales se le pide al alumno que identifique o reconozca la adecuación de un procedimiento o el uso adecuado de una herramienta; por ejemplo, en actividades de laboratorio se podrían presentar aparatos y herramientas, debiéndose indicar para qué y cómo se usan (v.g., verniers, balanzas de precisión, galvamómetros, voltímetro y otros).

2. Otras en las que se le solicita al sujeto que ejecute una tarea bajo condiciones simuladas en pequeña escala. En este caso, el alumno debe ejecutar movimientos, actividades y operaciones “como si” fueran reales, sea en modelos contruidos a escala o sin ellos.

3. Aquellas en los que se le pide al alumno que lleve a cabo una actividad muy representativa de una tarea o de una serie de tareas. En este caso si el sujeto sale airoso en la ejecución de dicha muestra de trabajo, se

podrá inferir que es capaz de resolver con eficacia todas las de su género. Este tipo de pruebas constityen criterios más directos y válidos de competencia que los exámenes de respuestas breves o de discusión, los cuales sólo pueden medir inferencialmente las mismas capacidades.

### **Interpretaciones de los Resultados de las Pruebas**

Existen tres maneras diferentes de interpretar las puntuaciones de las pruebas de rendimiento académico. Estas son:

1. La evaluación intrasujeto. Esta consiste en juzgar el desempeño del estudiante en la prueba tomando como criterio su propio nivel de capacidad, determinado por su ejecución previa en algún examen diagnóstico (prueba de aptitud, de conocimientos previos u otro similar). Este tipo de evaluación es importante porque permite conocer, tanto al alumno como al docente, el nivel de progreso de aquél a través del desarrollo del curso.

2. La evaluación normativa. En este caso se juzga el desempeño de un estudiante, tomando como criterio de referencia una norma de grupo o de tipo regional o nacional. Las medidas con referencia a una norma permiten comparar a los individuos con sus compañeros reales en la situación de aula de clase, o con un grupo mayor de compañeros, como podría ser una muestra nacional.

3. La evaluación referida a criterio. En algunos casos lo que se requiere es una norma absoluta de desempeño, no relativa, como en los casos anteriores, que sea independiente de la ejecución de otros o de la posición relativa del individuo dentro del grupo; es decir, cuando se desea conocer el dominio real que un individuo tiene de un campo.

Por ejemplo, en el caso de un estudiante de medicina, en un examen de cirugía, lo que interesa saber, desde el punto de vista del propósito de la

evaluación, es cuan eficiente es el sujeto haciendo una intervención quirúrgica, y no tanto, cómo es su progreso individual o cómo es su desempeño con respecto a su grupo, ya que podría darse el caso de que el sujeto haya progresado significativamente con respecto a su nivel inicial, o estar muy bien con respecto a la norma de su grupo y, sin embargo, no tener el dominio total de un determinado conocimiento o la competencia profesional requerida. Glaser (1963) llama a tales puntuaciones, basadas en una norma absoluta de calidad, medidas con referencia a un criterio, en contraste con las medidas con referencia a una norma. En otras palabras, una medida con referencia a un criterio revela lo que un individuo puede hacer, en lugar de indicar qué posición ocupa en comparación con su grupo de compañeros (Pophan y Husek, 1969).

# Pruebas Formales e Informales de Rendimiento Académico

Como se mencionó en el capítulo anterior, desde la creación de la Escuela como institución responsable de la administración de la educación formal, se idearon diferentes procedimientos o pruebas para evaluar el nivel de conocimientos obtenidos por los alumnos antes, durante o al final del proceso instruccional. Las pruebas escolares suelen cumplir diferentes propósitos; así permiten:

1. Conocer cuáles son las fortalezas y debilidades de los alumnos, previo al inicio del estudio de la asignatura, a fin de adecuar la misma al perfil de entrada del estudiante.
2. Determinar el nivel de progreso de los alumnos a través del curso a fin de darles retroalimentación oportuna y para reorientar la enseñanza en caso de que sea necesario.
3. Calificar a los alumnos e informar a los padres y representantes.
4. Consolidar el conocimiento adquirido, cuando se utilizan como guía para orientar el estudio y la preparación de exámenes futuros.

Adams (1970) considera que el maestro suele formularse ciertas preguntas claves durante el proceso de planificar, administrar e interpretar los resultados de una prueba de rendimiento, tales como: ¿Qué tipos de conductas pueden servir de prueba para afirmar la existencia de un progreso con relación a cada uno de los objetivos? ¿Qué clase de situación y qué

preguntas concretas pueden poner de manifiesto conductas de este tipo? ¿Cómo calificar o puntuar estas conductas? ¿Es este objetivo realista? ¿Cómo reorientar la enseñanza para obtener mejores resultados? Las pruebas de rendimiento escolar pueden ser clasificadas en informales y formales, según sea el nivel de sofisticación técnica y científica que se utilice en su proceso de elaboración y validación.

### **Las Pruebas Hechas por el Docente**

Tradicionalmente se han identificado con el nombre de pruebas informales a aquellas que elabora el docente para cumplir con su función evaluadora del rendimiento escolar o académico. Esta denominación se debe a que, por lo general, los docentes elaboran tales pruebas sin apearse mucho a los procedimientos técnicos establecidos para la elaboración de los instrumentos de medición. De allí la dificultad que normalmente se presenta para interpretar los resultados de tales pruebas.

Para que los procedimientos tradicionales de evaluación del rendimiento escolar tengan una mayor utilidad práctica, el proceso de su construcción debería aproximarse, lo más posible, al de las pruebas formales. En este sentido, recomendamos seguir los pasos siguientes:

1. Seleccionar (o definir) los objetivos objeto de la evaluación. Al respecto, el docente deberá considerar si se trata, por ejemplo, de objetivos del dominio cognoscitivo, afectivo o psicomotor y cuál será el nivel más apropiado de los objetivos. En el caso del dominio cognoscitivo, el docente deberá optar entre objetivos a nivel de conocimiento, comprensión, aplicación, análisis, síntesis o evaluación. Además de los objetivos, el docente deberá considerar, también los contenidos de la asignatura objeto de evaluación.

2. Determinar el tipo de procedimiento que puede ser más apropiado

a los fines que se persiguen con la evaluación. En este sentido, habrá que tomar en cuenta los sujetos (por ejemplo, edad y nivel educativo) a ser evaluados y la naturaleza de los objetivos instruccionales. Al respecto, el docente deberá decidir entre el uso de pruebas escritas, orales o prácticas. Y, en el caso de ser escritas, decidirá si utiliza pruebas objetivas o de ensayo. Todavía, si opta por la modalidad de pruebas objetivas, deberá decidir acerca del tipo de reactivo que puede ser más conveniente emplear (ítemes de selección simple o múltiple, falso o verdadero, pareamiento o completación).

3. Elaborar una Tabla de Especificaciones, la cual consiste en una matriz de planificación, organizada con base en dos variables: (a) los objetivos a ser evaluados y sus niveles correspondientes; y (b) los contenidos de la unidad curricular a ser evaluada. En las celdas que quedan definidas entre las columnas (objetivos) y las filas (contenidos) de la matriz, se indicará el número de ítemes o reactivos a ser utilizados (ver Cuadro 25).

*Cuadro 25. Tabla de especificaciones*

Contenidos	Objetivos						Total
	Conoc.	Comp.	Aplic.	Anál.	Sínt.	Eval.	
Unidad 1	2	1	3	1			7
Unidad 2			2	2	1		5
Unidad 3		2	1	1			4
Unidad 4	2	3	3				8
Unidad 5	3	3	2	4			12
<b>Total</b>	<b>7</b>	<b>9</b>	<b>11</b>	<b>8</b>	<b>1</b>		<b>36</b>

Como se puede observar en Cuadro , la Tabla de Especificaciones indica que la prueba objeto de planificación tendrá un total de 36 ítemes, de los cuales 7 corresponderán al nivel de conocimiento, 9 al de comprensión, 11 al de aplicación, 8 al de análisis y 1 al nivel de síntesis. Por otra parte, se puede apreciar que de los 36 ítemes de la tabla, 7 corresponden a la unidad uno; 5 a la dos; 4 a la tres; 8 a la cuatro y 12 a la unidad cinco.

4. Elaboración de los ítemes e integración de la prueba. Al respecto se deberá atender a las características de claridad en la redacción, a la pertinencia o congruencia ítem-dominio (contenido) y el control del sesgo en la formulación de los reactivos. También se deberá prestar atención a la claridad de las instrucciones y a la organización más apropiada de la información (ítemes) en el papel.

5. Prueba o revisión del instrumento. En este sentido, es recomendable utilizar la ayuda de otros colegas, especialistas en la misma asignatura, a los fines de determinar la validez de contenido de la prueba. En caso, de existir un Departamento o Unidad de Evaluación en la institución, no se debe dudar en acudir al mismo en solicitud de orientación al respecto.

6. Elaboración de modelos o claves de las respuestas, lo cual facilitará la corrección de las pruebas y permitirá controlar, en una buena medida, el sesgo o subjetividad en la asignación de las puntuaciones correspondientes.

7. Información a los alumnos acerca del tipo de prueba, condiciones, propósitos, objetivos y contenido a ser evaluados.

***Interpretación de los resultados.*** Existen diferentes criterios para interpretar los resultados (calificaciones) de las pruebas escolares. Dichos criterios son establecidos por las autoridades competentes, a objeto de darle la

validez legal correspondiente. En el caso de Venezuela, en los niveles de Educación Preescolar, Básica y; Media, Diversificada y Profesional, le corresponde al Ministerio de Educación, a través de la Dirección General Sectorial de Apoyo Docente establecer la normativa que rige la evaluación escolar en el ámbito del territorio nacional.

Un criterio vigente (escala 1 a 20) para interpretar las calificaciones escolares con un propósito evaluativo, es el de utilizar el puntaje 10 para decidir quiénes están aprobados (calificación igual o mayor que 10) o reprobados (calificación igual o menor que 9). Otro criterio, consiste en utilizar ciertas categorías evaluativas que hacen referencia al desempeño del estudiante en la prueba, en función de los puntajes obtenidos, como se indica más abajo.

<b>Puntuaciones</b>	<b>Categorías</b>
18 a 20	Excelente
15 a 17	Bueno
13 a 14	Satisfactorio
10 a 12	Regular
01 a 09	Deficiente

Esta manera de interpretar los resultados de las pruebas tiene sentido cuando el propósito de la evaluación es promover al estudiante de un grado a otro. Pero no es tan útil cuando lo que interesa es, por ejemplo, diagnosticar el nivel de conocimientos de los alumnos. Ilustremos con un ejemplo. Supongamos que se ha administrado una prueba de 10 reactivos y dos alumnos obtienen las calificaciones de 5 puntos cada uno. Si hacemos una transformación lineal, mediante una regla de tres, a la escala de 1 a 20, estos

estudiantes tendrían una calificación de 10 puntos; es decir, su desempeño sería evaluado como regular.

Tal calificación indica que estos alumnos presentan algunas deficiencias en cuanto al conocimiento que deberían tener. Sin embargo, el hecho de haber obtenido la misma calificación no es un indicador válido de que los dos tienen los mismos conocimientos de la asignatura, ni mucho menos que tienen las mismas deficiencias. Ello se explica porque una calificación de 5 puntos en una prueba de 10 reactivos puede ser obtenida con diferentes combinaciones de ítems respondidos correctamente, los cuales no tienen que ser necesariamente los mismos en cada uno de los dos casos. Entonces, ¿cómo se podría saber cuáles son las necesidades de cada uno de ellos? Una manera de responder esta pregunta es elaborando una matriz de objetivos/item/sujeto, en la cual se pueda apreciar el número de ítems que cada alumno respondió correctamente para obtener una determinada calificación y, sobre todo, cuáles ítems los estudiantes no respondieron o lo hicieron erróneamente (ver Cuadro 26).

*Cuadro 26. Matriz de resultados de una prueba*

Ítemes	Objetivo 1			Objetivo 2			Objetivo 3				Total
	1	2	3	4	5	6	7	8	9	10	
Alumnos											
A	x	x	✓	✓	✓	x	x	✓	✓	x	5
B	✓	✓	x	✓	x	✓	x	x	x	✓	5
C	x	x	x	x	✓	x	✓	x	✓	✓	4
D	x	x	x	x	✓	x	✓	✓	✓	✓	5
E	x	x	✓	✓	✓	✓	✓	✓	✓	✓	8
Total	1	1	2	3	4	2	3	3	4	4	

✓ = Correcta      x = Incorrecta

Como se puede observar en el Cuadro 26 , los estudiantes A y B obtuvieron 5 como calificación en la prueba. Sin embargo, se puede apreciar que la puntuación del alumno A está conformada por los ítemes: 3, 4, 5, 8 y 9; mientras que la nota del estudiante B está compuesta por los reactivos: 1, 2, 4, 6, y 10. Tal resultado indica que el estudiante A respondió una pregunta en el objetivo 1 y 2 en los objetivos dos y tres; mientras que el alumno B contestó dos preguntas en el objetivo 1, dos en el objetivo 2 y una en el objetivo 3. Todo lo cual indica que estos dos estudiantes tienen conocimientos diferentes del área de contenido y, también presentan necesidades diferentes que deben ser atendidas, a pesar de haber obtenido la misma calificación en la prueba. De allí que se recomiende utilizar, en las pruebas diagnóstica, un análisis de tipo cualitativo previo, con base en la matriz antes señalada, para poder interpretar el significado de las calificaciones en el caso de cada estudiante.

Los educadores han utilizado tanto las formas tradicionales de evaluación centradas en las pruebas escritas, orales y prácticas, como las

pruebas de nuevo tipo u objetivas ideadas posteriormente. Brownell (citado por Adams, 1970) observa que la preocupación esencial del maestro, al desarrollar sus propias pruebas, debe ser la de mejorar la calidad de la enseñanza mediante la evaluación. En tal sentido, cree que dichas pruebas deben: (a) promover el desarrollo de procesos cognitivos en los estudiantes; (b) estimular a los alumnos para estudiar relaciones y aplicar principios; y (c) fomentar el desarrollo de hábitos de estudio deseables.

### **Las Pruebas Formales**

Además de las pruebas elaboradas por los docentes (pruebas informales), el avance de la tecnología de la medición y la evaluación han permitido a los especialistas de este campo utilizar procedimientos formales, cada vez más sofisticados, para cumplir con el propósito de evaluar el rendimiento estudiantil. Surgieron así, primero, las pruebas basadas en normas y, posteriormente, las pruebas referidas a criterio.

***Pruebas basadas en normas.*** Las pruebas de rendimiento académico basadas en normas (PBN) son instrumentos que permiten determinar el nivel de conocimientos alcanzado por los alumnos (individualmente o en grupo), comparando el desempeño de éstos en la prueba con los valores normativos, obtenidos previamente en un grupo de referencia. Es decir, “una prueba basada en normas es diseñada con el propósito de precisar el estatus de un examinando en relación con la ejecución de un grupo de otros examinandos quienes han completado dicha prueba” (Popham, 1978, p. 24). Para construir este tipo de instrumento, el especialista procede conforme a los siguientes pasos generales:

1. Se planifica y construye la prueba con una extensión determinada, de acuerdo con la necesidad a satisfacer y la naturaleza del área de contenido de que se trate. Digamos, por ejemplo, que se ha elaborado una

prueba de 60 ítemes para evaluar el dominio de la Lengua Materna (Castellano) en el segundo nivel de Educación Básica, del Estado Lara.

2. El instrumento se administra previamente en una muestra grande (por ejemplo,  $n = 1000$  sujetos) y representativa de alumnos; es decir, en sujetos que, además de pertenecer al mismo nivel educativo, tengan características similares a las de los alumnos con quienes finalmente se utilizará la prueba.

3. Se corrigen y puntúan las pruebas; se realiza el estudio técnico correspondiente (análisis de ítem, confiabilidad y validez de la medida) y luego se calculan las normas de la misma. Estas normas se suelen expresar en términos de: (a) medidas de tendencia central como la media; y de variabilidad, como la desviación típica; (b) medidas de posición como los percentiles, deciles o cuartiles; y (c) medidas estandarizadas como las expresadas en la escala T o en estaninas. Retornando al ejemplo anterior, supongamos que la media aritmética y la desviación típica de la ejecución de la muestra de alumnos ( $n = 1000$ ) fue de  $M = 38$  y 6, puntos respectivamente y que los percentiles 25, 50 y 75 fueron 26, 38 y 50 puntos.

4. Se administra el instrumento en un grupo particular; por ejemplo, un grupo de 120 alumnos de la población de Cabudare (Municipio Autónomo Palavecino) de segunda etapa de Educación Básica. Suponga, además, que la distribución de puntajes de los alumnos fue como se observa en el Cuadro 27.

*Cuadro 27. Distribución de puntajes en la prueba de lengua materna en una muestra de alumnos de cabudare (Edo. Lara).*

Puntajes	Alumnos	Porcentajes
25 y menos	48	40
26 a 37	36	30
38	18	15
39 a 49	12	10
50 y más	6	5
	120	100%

Los resultados anteriores indican que apenas 36 alumnos (30 por ciento) del total ( $n = 120$ ) obtuvo puntuaciones igual o superior a la norma (Media = 38); mientras que 84 estudiantes (70 por ciento) no alcanzaron el nivel de desempeño esperado (la media aritmética de las calificaciones del grupo).

Un análisis similar se podría hacer tomando el percentil 50 (o la Mediana) como norma. De estos resultados ficticios se podría concluir que la mayoría de los alumnos de la muestra no tiene un dominio apropiado de la Lengua Materna, de acuerdo con el nivel exigido en la segunda etapa de Educación Básica en el Estado Lara (Venezuela).

Las PBN llegaron a ser muy populares en países como los Estados Unidos de Norte América, hasta la década de los años 70, particularmente debido al extraordinario desarrollo alcanzado en la psicometría, el auge de los paquetes estadísticos computarizados, el dominio del conductismo como teoría psicológica y su influencia en la instrucción, expresado a través del uso de las máquinas de enseñar, la instrucción programada, la instrucción personalizada, tipo Plan Keller, y el uso de objetivos instruccionales en términos de conducta observable.

Sin embargo, pronto algunos educadores y evaluadores se dieron cuenta que mientras estas PBN parecían ser muy útiles en los casos de selección de

alumnos para determinados fines; eran poco apropiadas para ser usadas como instrumento de evaluación del trabajo escolar. Al respecto, algunos autores (Bloom, 1976; Popham, 1978; Nunnally (1972) han considerado que muchas de estas pruebas resultan ser poco válidas para propósitos específicos de evaluación del aprovechamiento de los alumnos, ya que sólo nos ofrecen una visión relativa del desempeño académico, al comparar a los alumnos con respecto a una norma de referencia, y no permiten identificar qué áreas específicas de contenido dominan los alumnos y en cuáles necesitan mejorar.

En el caso específico de Venezuela, no ha existido una tradición en el uso de este tipo de pruebas de rendimiento basadas en normas, en el contexto de la evaluación escolar. Sin embargo, se concocen algunas experiencias al respecto que vale la pena mencionar. Por ejemplo, el Ministerio de Educación (1970) diseñó un programa de evaluación de rendimiento estudiantil en Matemáticas y Lenguaje en el nivel de sexto grado, para lo cual construyeron batería de pruebas estandarizadas.

Estudios similares han sido realizados por la Oficina de Planificación del Sector Universitario (OPSU) y el Centro Nacional para el Mejoramiento de la Ciencia (CENAMEC), en relación con la evaluación de los conocimientos de los alumnos en el Ciclo Básico Común, hoy llamado tercera etapa de Educación Básica (ver OPSU-CENAMEC, 1987). Al mismo tiempo, la OPSU ha venido empleando desde el año de 1982 este tipo de instrumento (como es el caso de la Prueba de Aptitud Académica) a nivel nacional, como requisito de ingreso al subsistema de educación superior en Venezuela (ver Silva, De Sanctis y Saarco, 1982). Las PBN se diferencian de las pruebas hechas por el docente de aula (PHD), en los siguientes aspectos:

1. **Cantidad de contenido.** La prueba estandarizada cubre mayor cantidad de contenido que la PHD. Ello se debe a que en esta última el

contenido, por lo general, se refiere a la evaluación de la materia vista durante un período determinado de tiempo (v. g., una semana, un mes, un trimestre, un semestre o un año escolar); mientras que las PBN están más bien referidas a un muestreo de ítemes basado en la totalidad del contenido que cubre una asignatura en un nivel determinado (por ejemplo, matemática de noveno grado de Educación Básica).

2. **Propósito.** La prueba PHD, tiene un carácter situacional, ya que intenta medir el rendimiento en una asignatura administrada por un docente dentro de una institución escolar o programa de estudio. Las PBN intentan medir el nivel de rendimiento que los estudiantes han alcanzado en una materia determinada en el contexto: (a) institucional (por ejemplo, matemática de las diferentes secciones de sexto grado en la Unidad Educativa “Luis Beltrán Prieto Figueroa” de Barquisimeto, Venezuela); (b) local (ejemplo, matemática de sexto grado en el Distrito Iribarren de Barquisimeto); (c) regional (por ejemplo, matemática de sexto grado en los Estados Yaracuy, Lara y Falcón); y (d) nacional (por ejemplo, matemática de sexto grado de todas las escuelas del país).

3. **Construcción.** Las PHD requieren de menos sofisticación técnica para su construcción que las PBN. Además, las pruebas escolares requieren de menos tiempo para su elaboración y su costo, por este concepto, es también menor.

4. **Normas.** Una de las ventajas más importantes de las pruebas estandarizadas es que ellas proveen al evaluador de normas que permiten comparar el rendimiento de los estudiantes de manera individual, como grupo de una clase o como plantel escolar, con respecto a un grupo normativo de carácter local, regional o nacional. Por ejemplo, si sabemos que el promedio en matemática en el nivel de Educación Media es de 13

puntos (norma nacional obtenida para una prueba estandarizada) y que el promedio obtenido en la misma asignatura y nivel educativo, para el liceo “Lisandro Alvarado” de Barquisimeto fue de 11 puntos, podemos darnos cuenta, que los alumnos de dicho liceo tienen un rendimiento en matemática inferior a la norma nacional.

Los docentes de aulas no utilizan pruebas estandarizadas para evaluar a sus alumnos, sino que emplean instrumentos de evaluación escolar que a lo sumo tienen validez de contenido y, por lo general, se califican con relación a los valores de la escala de calificaciones oficial (por ejemplo, la escala de 1 a 20).

5. **Uso.** La comparación que se puede hacer de estos dos tipos de pruebas cae en el campo de la validez. Ambos tipos de instrumentos intentan medir el rendimiento escolar, pero con un propósito diferente. La PHD es válida para medir cuánto sabe el estudiante con respecto a los contenidos que el docente ha impartido, durante un período de tiempo determinado; y a lo sumo permite hacer comparaciones del rendimiento individual de los alumnos con respecto al desempeño del grupo, o con grupos de clase de años o semestres anteriores, en el caso de que se utilice la misma prueba.

La prueba estandarizada permite más hacer comparaciones con respecto a normas de grupo que han sido obtenidas mediante procedimientos técnicos sofisticados y laboriosos. De esta manera, permite obtener diagnósticos válidos del rendimiento escolar por asignatura y nivel educativo en el contexto nacional, regional o estatal.

6. **Administración.** Las pruebas estandarizadas se administran siempre de la misma manera en cuanto a las instrucciones de la prueba, el uso del tiempo y de las condiciones ambientales para su aplicación. En el

caso de las PHD, la administración de estos instrumentos se basa en reglas más sencillas y flexibles; es decir, su administración es menos rigurosa y estructurada.

***Pruebas referidas a criterio.*** La literatura especializada atribuye a Glaser (1963) la distinción entre: (a) “prueba basada en normas”, que él consideraba centrada en la detección del puesto relativo de un estudiante, en relación con un grupo de referencia; y (b) “prueba referida a criterio” (PRC), que él creía estaba centrada en un nivel absoluto de calidad. Glaser llamó la atención a educadores y psicólogos sobre lo poco adecuado que era aplicar conceptos tradicionales de medición a situaciones en las que se ponían en juego estrategias innovadoras de enseñanza.

Es bien conocido, en el campo de la medición y evaluación, que la aptitud académica tiende a distribuirse conforme al modelo de la curva normal de probabilidades, lo cual permite esperar una distribución similar del rendimiento escolar. Es decir, bajo condiciones estándares de instrucción (una misma modalidad de enseñanza para todos los alumnos), es de esperarse que: (a) los alumnos que son bajos en aptitud al comienzo de la instrucción, obtengan bajas calificaciones al final del curso; (b) aquellos quienes tenían aptitud alta al comienzo, obtendrían calificaciones altas al final; y (c) los alumnos con aptitud promedio inicialmente, obtendrían calificaciones promedio al final del curso. Todo lo cual puede expresarse afirmando que es esperable una alta correlación positiva y significativa entre la aptitud académica de los estudiantes y su rendimiento escolar.

Por otra parte, se asume que bajo condiciones especiales de instrucción, como en el caso en que se utilizan estrategias innovadoras de enseñanza, la mayoría de los alumnos (alrededor del 90 por ciento) deben tener éxito; es decir, deben alcanzar un dominio máximo de los objetivos instruccionales. Si

esto es así, la distribución de los puntajes del rendimiento no sería normal sino asimétrica negativa; o sea que la mayoría de las puntuaciones estarían concentradas hacia el extremo derecho de la curva y su relación con la aptitud inicial sería baja con tendencia a cero, dependiendo de la efectividad de la instrucción.

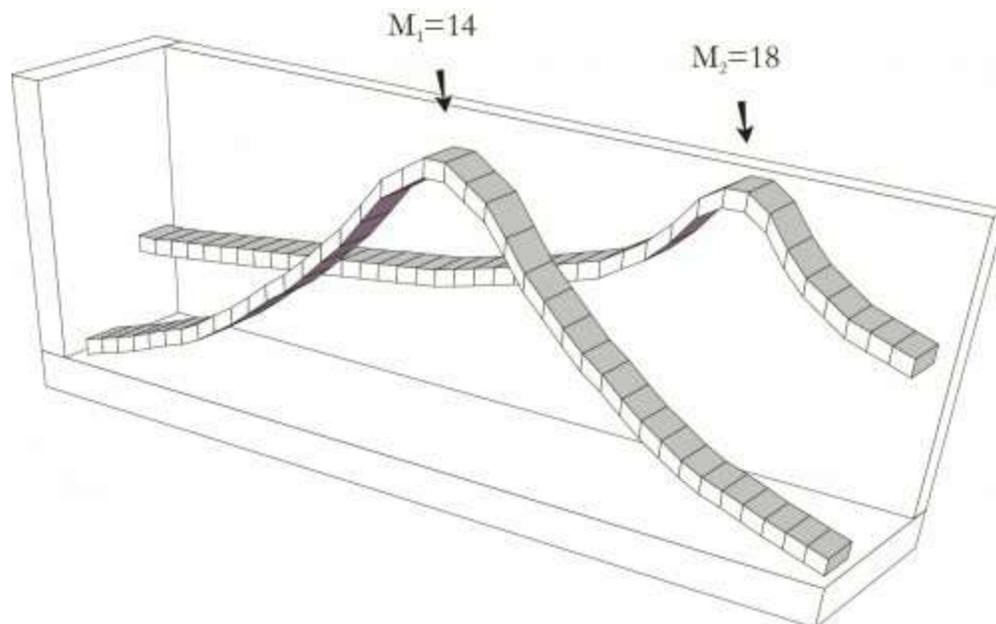
Este tipo de relación entre la aptitud inicial, el rendimiento académico y la calidad de la instrucción ha sido apoyada con datos experimentales por investigadores como Glaser (1963), Bloom (1976) y Pophan (1980), entre otros. Con el propósito de ayudar al lector a visualizar mejor las relaciones entre estas variables, se ha diseñado un ejemplo con datos ficticios, los cuales se presentan en el Cuadro 28, complementado con la Figura 4, basada en dichos datos.

*Cuadro 28. Aptitud y rendimiento académico medido con una prueba basada en normas (PBN) y con una prueba referida a criterio (PRC)*

Sujetos	Aptitud (1)	Rendimiento	
		PBN (2)	PRC (3)
1	80	20	20
2	59	18	20
3	68	17	19
4	64	16	19
5	64	16	18
6	70	15	18
7	60	15	18
8	64	16	18
9	64	16	17
10	70	15	17
11	60	14	17
12	56	14	17
13	52	13	20
14	52	13	20
15	56	14	16
16	56	14	19
17	52	13	20
18	54	12	20
19	44	11	20
20	40	10	20

Sin embargo, este tipo de resultados no es posible cuando se utilizan las PBN, ya que los criterios utilizados en su construcción no lo permiten. Por ejemplo, en el proceso de construcción de una PBN, un ítem que es repondido satisfactoriamente por todo el grupo de alumnos es considerado

como muy fácil y sin ningún poder de discriminación; por lo tanto, el mismo no puede ser seleccionado para formar parte de la versión final del instrumento; mientras que en una prueba referida a criterio, este tipo de ítem es incluido en la versión final de la prueba, por cuanto es un buen indicador de la efectividad del docente y del aprovechamiento de los alumnos. De allí que se pusiera un poco en duda la validez de las pruebas basadas en normas, como instrumentos de evaluación del rendimiento escolar, y se prefiriese las pruebas referidas a criterio.



*Figura 4. Dos distribuciones de calificaciones con una PBN y con PRC*

Pero, ¿qué se entiende por criterio en este contexto? Esta palabra creó cierta confusión al comienzo, ya que no fue interpretada de manera similar por todos los educadores y evaluadores. Para algunos, la palabra **criterio** fue identificada con **objetivo de instrucción**, particularmente por la popularidad que llegaron a tener los objetivos educacionales durante las décadas de los años 60 y 70. De allí que al hablar de pruebas referidas a criterio se

entendiera que era un tipo de instrumento de evaluación centrado en la descripción de los objetivos de conducta.

Para otros, la palabra **criterio** fue identificada con el **nivel deseado de capacitación del alumno**, expresado como el dominio mínimo que se espera del sujeto en una prueba de rendimiento, en relación con una unidad de aprendizaje, como ocurre, por ejemplo, en el “mastery learning” (ver Bloom, 1976) o en el Plan Keller (Keller, 1968). De allí que cuando se empezó a hablar de pruebas referidas a criterio, algunos entendieron que se trataba del puntaje mínimo requerido para aprobar una unidad instruccional.

Finalmente, para otros, la palabra **criterio** estuvo asociada con un **resultado postinstruccional** deseado que permitiera determinar la validez del proceso instruccional de acuerdo con el desempeño del sujeto en situaciones de la vida diaria; por ejemplo, la enseñanza de la aritmética en Educación Básica (segunda etapa) y el manejo del niño en un supermercado, o la enseñanza de la Educación Moral y Cívica y el comportamiento ciudadano de los alumnos, una vez que están fuera del ámbito escolar. De allí que algunos entendieran que las pruebas referidas a criterio estaban relacionadas con este tipo de situaciones de la vida real, que permiten al sujeto probar la validez de los resultados de su aprendizaje escolar frente a los retos de la vida diaria.

No obstante lo anterior, la idea que se quiso expresar con la palabra criterio fue la de un **dominio de conducta bien definido**. Un dominio de conducta consiste en un conjunto de habilidades y disposiciones que el sujeto exhibe cuando se le pide hacerlo en el contexto de una situación de evaluación. Por ejemplo, un dominio de conducta en el campo de la estadística podría consistir en la habilidad de un individuo para resolver cierta clase de problemas con base en el modelo de la curva normal de

probabilidades. Un dominio de conducta en el campo de las actitudes podría estar constituido por los tipos de respuestas que un individuo podría dar a un conjunto de ítemes de escogencia forzada, en los que se refleja la preferencia del sujeto por ciertas actividades de ocio.

Un dominio de conducta puede ser definido de tal manera que exprese una realidad de naturaleza compleja, la cual podría consistir en un conjunto de habilidades de alto nivel que, a su vez, incluyan a otro conjunto de habilidades de nivel más bajo. Por otra parte, un dominio de conducta puede ser definido de tal manera que aparezca como algo realmente minúsculo; por ejemplo, evaluar la habilidad del sujeto para escribir correctamente 10 palabras.

La situación anterior, introduce una dificultad importante, en el proceso de elaborar pruebas referidas a criterio, ya que no se han establecido límites exactos que permitan precisar lo que es un dominio bien definido. En general, muchos autores están de acuerdo en que un dominio bien definido es aquel que presenta una descripción lo más detallada posible de dicho dominio conductual objeto de la evaluación. Debido a que una buena prueba referida a criterio está formada por un conjunto de reactivos derivados de un dominio específico de conducta, algunos autores han escogido tal expresión (prueba referida a criterio) para designar a este tipo de instrumento de evaluación del rendimiento escolar.

El propósito principal de las PRC consiste en presentar el marco conductual con base en el cual el desempeño del sujeto va a ser evaluado. De allí que la tarea más importantes de los diseñadores de este tipo de instrumento de evaluación consista en diseñar mecanismos efectivos que permitan establecer exactamente cuál es el significado real del desempeño del sujeto. Pero como existen diferentes maneras de describir tal desempeño en la

prueba, se ha decidido incluir todas estas posibilidades bajo la categoría denominada **descripción de campo o esquema descriptivo**.

## Descripción de Campo

La descripción del campo representa el conjunto de reglas necesarias para generar los elementos de la prueba, lo cual constituye el paso más difícil y más importante en la construcción de PRC que sean válidas. Estas descripciones deben tener las siguientes características:

1. Deben ser lo suficientemente breve para resultar utilizable, no sólo por los encargados de elaborar los reactivos, que son quienes tienen que guiarse por ellas, sino también por los educadores quienes pueden emprender una acción determinada con base en los datos de la prueba.

2. Deben circunscribir suficientemente la clase de conductas que se estudia, de manera que distintos observadores independientes alcancen un alto grado de coincidencia en el dictamen sobre si determinados elementos de la prueba miden realmente la conducta descrita en ese campo.

El esquema descriptivo de la prueba representa un intento, del diseñador del instrumento, por comunicar a terceras personas, cuál es el dominio de conductas que pretende medir la prueba. Típicamente, el esquema descriptivo de la prueba será de naturaleza exclusivamente verbal, aun cuando ciertas clases de conductas, tales como las psicomotoras, el esquema descriptivo podrían incluir elementos pictóricos y simbólicos.

El esquema descriptivo cumple dos funciones básicas. En primer lugar, intenta comunicar al usuario de la prueba qué es lo que dicho instrumento mide realmente y, en segundo lugar, busca comunicar a los diseñadores de este tipo de pruebas qué clase de reactivos son los más apropiados para ser incluidos en una PRC en particular. De allí que algunos autores se refieran a

todo este tipo de descripciones como **especificaciones de la prueba**.

Existe una íntima relación entre el propósito del esquema descriptivo del dominio conductual, la elaboración de los ítemes por parte del diseñador de la prueba y el usuario del instrumento. En tal sentido, es de esperarse que un buen esquema descriptivo, permita generar reactivos congruentes con el dominio conductual y hacer interpretaciones confiables de los resultados de la prueba, por parte de los usuarios. Y, a su vez, un esquema descriptivo vago o impreciso, conduce a una elaboración de reactivos irrelevantes o no congruentes, todo lo cual producirá unos resultados de difícil interpretación.

### **Especificaciones de la Prueba**

Como ya se mencionó anteriormente, existen dos funciones básicas que cumplen las especificaciones de una PRC: (a) comunicar a los usuarios qué es lo que la prueba mide, a fin de que éstos puedan comprender con seguridad la naturaleza de la conducta que se está evaluando; y (b) presentar descripciones detalladas del dominio conductual que se pretende medir, a fin de que el diseñador de la prueba pueda generar un conjunto de reactivos homogéneos y funcionales. Es decir, aquellos ítemes que miden el estatus de sujeto con respecto al dominio de conducta descrito en las especificaciones de la prueba. En otras palabras, el propósito de las especificaciones de la prueba es comunicar, en primer lugar, al diseñador y, posteriormente, a los usuarios. Existen diferentes maneras de crear especificaciones funcionales para la elaboración de una PRC. No obstante, en este capítulo señalaremos una secuencia de pasos al respecto sugerida por Popham (1978).

**Tener presente qué medirá la prueba.** La primera etapa, en el establecimiento de las especificaciones de la prueba es tener en mente una idea general acerca de lo que la prueba va a medir. Por ejemplo, en una prueba de tipo cognitiva o psicomotora, se intenta saber qué es lo que el

sujeto es capaz de hacer; mientras que en el caso de una prueba del área afectiva se aspira a determinar cuáles son las disposiciones del alumno.

**Nivel de generalidad de la conducta.** En segundo lugar, es necesario determinar el nivel de generalidad de la conducta que se pretende medir. Al respecto, se recomienda utilizar pequeños, pero altamente importantes, segmentos de la conducta del sujeto.

Esto facilita el que se pueda aislar más fácilmente y describir las dimensiones claves de la conducta a ser medida. Las conductas complejas son más difíciles de definir con suficiente especificidad.

**Selección de alternativas.** La tercera etapa en el establecimiento de las especificaciones de la prueba, consiste en seleccionar las conductas a ser medidas. En esta etapa se requiere elegir uno de los posibles campos que pueden ser utilizados como criterio para medir la conducta. Por lo general, se elige uno (entre varios, por ejemplo, conocimientos, actitudes, habilidades, estrategias) que juzgamos relevante a tal fin. Pero para muchas clases de conducta existen varios otros posibles campos que podrían emplearse legítimamente para representar la posesión de la conducta propuesta como meta.

Se debe seleccionar la forma de conducta del alumno que sea más generalizable, entre las posibles alternativas; o sea, aquella que mejor garantice, que una vez adquirida ésta, pueda ser transferida, a la mayor cantidad de campos posibles, tanto en los eventos relacionados con tareas académicas, como con aquellos que se refieren a la vida diaria. Es decir, es importante tomar en cuenta, en el momento de seleccionar el tipo de campo específico, el grado en que éste, cuando se posee, pueda ser transferido a otros campos con los cuales se relacionará el sujeto en el presente o en el futuro.

## **Elementos de una Descripción de Campo**

Existen muchas maneras como se pueden establecer las especificaciones de una PRC. En este capítulo se utiliza un esquema que ha sido suficientemente ensayado por Popham (1978), el cual incluye los siguientes elementos: descripción general, ítem de muestra, límites del contenido, límites de las respuestas, criterios de corrección y especificaciones complementarias. A continuación se describe cada uno de estos elementos.

**Descripción general.** Constituye el primer componente de las especificaciones de una PRC. Consisten en una presentación breve y precisa acerca de lo que la prueba mide. Su propósito es proveer al evaluador de una visión global del conjunto de conductas que serán descritas más completamente posteriormente en las especificaciones. En algunas PRC este componente es equivalente a un objetivo, como se ilustra en el siguiente ejemplo: Caracterizar una distribución de puntajes tomando en cuenta uno o más de los estadísticos de tendencia central, variabilidad y posición.

**Ítem de muestra.** El segundo componente en las especificaciones de una PRC es la presentación del “ítem de muestra”. Como su nombre lo indica, se trata de incluir, un reactivo modelo conjuntamente con sus instrucciones, a fin de que le sirva de guía al diseñador del instrumento, para la redacción de los ítems de la prueba y, al mismo tiempo, sirva de orientación a los sujetos para que se familiaricen con el tipo de ítems o preguntas incluidos en la prueba, como se ilustra con el ejemplo que aparece más abajo:

*Ejemplo 14*

**Instrucciones:** A continuación se presenta una serie de cuestiones para resolver. Lea cuidadosamente cada proposición y realice la actividad que se le solicita. Dada una distribución de calificaciones de 10 estudiantes del primer semestre de Psicología de una universidad venezolana, calcule e interprete las medidas de tendencia central, variabilidad y posición que considere relevante para describir la siguiente distribución: 10, 15, 20, 8, 10, 13, 18, 16, 15, 15.

**Límites del contenido.** En esta sección se deben señalar todos los factores que influyen positiva o negativamente la composición de un conjunto de ítems. Esto significa que se debe, en primer lugar, identificar cuáles son tales factores para, luego, describirlos de una manera sucinta y segura. Es decir, en esta sección se le indica al diseñador de la prueba cuáles son las dimensiones críticas a ser tomadas en cuenta a fin de redactar un conjunto de ítems que sean congruentes con el dominio previamente definido, como se indica en el ejemplo que expone a continuación:

*Ejemplo 15*

1. Los ítems serán proposiciones tendientes a describir una distribución de datos de acuerdo con las medidas de tendencia central, variabilidad y posición.
2. Los datos deberán estar referidos preferiblemente a variables psicoeducativas, tales como: rendimiento académico, aptitudes, inteligencia, actitudes, motivación, ansiedad, entre otras.
3. Cada ítem debe estar referido a, por lo menos, una de las medidas de tendencia central, variabilidad o posición.
4. En cada ítem se debe incluir el nombre de la variable bajo estudio, el nombre del estadístico objeto del cálculo y la fórmula del mismo.
5. Preferiblemente se deben emplear reactivos objetivos con cuatro alternativas de respuesta.
6. Cada ítem deberá ser independiente de los demás.
7. El conjunto de ítems debe estar redactado de tal manera que refleje diferentes niveles cognitivos y de dificultad.
8. La prueba deberá estar integrada por igual número de ítems con respecto a los diferentes tipos de medidas: tendencia central, variabilidad y posición.

**Límites de las respuestas.** En cada ítem, el estudiante seleccionará la respuesta entre las alternativas disponibles y justificará su escogencia, como se ilustra en el ejemplo que aparece más abajo.

*Ejemplo 16*

Determine el nivel de rendimiento de 10 estudiantes de psicología cuyas calificaciones fueron: 10, 15, 20, 8, 10, 13, 18, 16, 15, 15. Marque con una equis (x) la alternativa de respuesta que usted considere correcta.

- a.  13,5
- b.  14,0
- c.  14,4
- d.  14,9

**Razonamiento:** El promedio del rendimiento académico del grupo fue 14 puntos, el cual puede ser interpretado como satisfactorio, para una escala de 1 a 20 puntos.

**Criterios de corrección.** En esta sección se indicará cómo serán valoradas las respuestas emitidas por los sujetos evaluados. En el caso anterior, los criterios de corrección serán los siguientes:

1. Cada pregunta será valorada en un punto.
2. Cuando el estudiante no escriba el razonamiento de la respuesta seleccionada, sólo se le adjudicará medio punto.

**Especificaciones complementarias.** Esta sección es opcional y se debe incluir en ella cualquier información adicional que permita ampliar la descripción del campo, a fin de que los diseñadores de la prueba tengan una visión más completa acerca de su alcance.

### **Características Técnicas: Confiabilidad y Validez**

La noción de confiabilidad y validez de la medida tiende a ser la misma, independientemente que se trate de una PBN o de una PRC. No obstante ello, existen algunas diferencias importantes que vale la pena considerar. Como ya se ha mencionado, una de las contribuciones de las PRC está en que nos permiten tener una mayor seguridad en la descripción del desempeño académico del sujeto. Pero, aun en el caso de que el desempeño haya sido bien descrito todavía existe el problema de qué tan bueno ha sido la ejecución académica a objeto de que sea considerada aceptable.

**Confiabilidad.** Como ya ha sido expresado en capítulos anteriores, la confiabilidad se refiere a la estabilidad o consistencia con que la prueba mide lo que pretende medir. Típicamente, el nivel de confiabilidad de una prueba se expresa a través de un índice de correlación, el cual, a su vez, depende de la variación implícita entre las variables. Pero, es el caso que cuando se

utilizan las PRC el rango de variación de las puntuaciones de los sujetos tiende a ser pequeño, por lo cual no siempre la confiabilidad de la medida puede ser establecida mediante del procedimiento clásico del coeficiente de correlación ya mencionado. En estos casos de varianza reducida se requiere de la utilización de técnicas estadísticas alternativas, como por ejemplo los modelos no-paramétricos.

Tales modelos permiten obtener una buena estimación del grado en que el desempeño de los sujetos en dos medidas está relacionado. Un ejemplo de ello lo constituye el coeficiente Phi. Para utilizar este modelo, se obtiene la mediana de las medidas (primera y segunda medición) y se clasifican los datos en dos categorías: (a) puntajes por encima de la mediana ( $P > Md$ ); y (b) puntajes por de bajo de la mediana ( $P < Md$ ), como se indica en el ejemplo que se presenta a continuación. Suponga que se ha administrado una PRC en dos ocasiones a los mismos estudiantes ( $n = 100$ ) y utilizando el criterio de la mediana se obtiene la clasificación que se presenta en el Cuadro 29.

*Cuadro 29. Clasificación de dos distribuciones de datos con base en la mediana*

		Medición 1		
		P > Md	P < Md	
Medición 2	P > Md	45 (A)	5 (B)	50
	P < Md	8 (C)	42 (D)	50
		53	47	100

Si se aplica la fórmula del coeficiente Phi:

$$Phi = \frac{AD-BC}{\sqrt{(A+B)(B+D)(A+C)(B+D)}}$$

Sustituyendo, tenemos:

$$Phi = \frac{(45)(42)-(5)(8)}{\sqrt{(45+5)(5+42)(45+8)(5+42)}} = \frac{1890-40}{\sqrt{(50)(47)(53)(47)}}$$

$$Phi = \frac{1850}{5853850} = \frac{1850}{2419,47}$$

$$Phi = 0,76$$

El valor del coeficiente phi (0,76) obtenido indica que hay una correlación alta entre los puntajes de las dos distribuciones de calificaciones (medición 1 y medición 2), lo cual permite concluir que la prueba tiene una confiabilidad aceptable a los fines que persigue la PRC.

Un análisis alternativo consiste en emplear descripciones más gruesas como las que se obtienen mediante el porcentaje de consistencia de la decisión. En este sentido, el análisis de los datos se centra en el porcentaje de sujetos que fueron consistentes al quedar por encima de la mediana en la mediciones 1 y 2, y el porcentaje de aquellos quienes quedaron por de bajo de la mediana en ambas mediciones. En el primer caso, el porcentaje fue de 90, mientras que en el segundo fue de 84, para un porcentaje promedio de consistencia de 87.

Existen otros casos en que el procedimiento del porcentaje de consistencia de la decisión puede ser empleado. Por ejemplo, cuando utilizamos PRC, en las cuales se establece un nivel de desempeño mínimo (por ejemplo 85 por ciento) como indicador de dominio. Así, podríamos comparar el porcentaje de sujetos que fue consistente en alcanzar el nivel mínimo de desempeño establecido entre la primera y la segunda medición.

En segundo lugar, se compararía el porcentaje de sujetos que fueron consistentes en cuanto a no lograr alcanzar el nivel mínimo de desempeño.

En el caso de las PRC el tipo de confiabilidad que puede ser relevante en un momento dado, dependerá del propósito de la prueba. Por ejemplo, sabemos que las pruebas pueden ser utilizadas con el propósito de: (a) tomar decisiones acerca del nivel de progreso de los alumnos; (b) para seleccionar a los futuros participantes de un curso dado; o (c) para evaluar la efectividad de un determinado programa. La naturaleza de la alternativa de decisión influenciará el tipo de procedimiento de confiabilidad que se va a emplear (Popham, 1978).

En general, los especialistas en el diseño de las PRC recomiendan utilizar los procedimientos correlacionales clásicos a los fines de estimar la confiabilidad de una medida, siempre y cuando exista una cierta variación en los datos. De lo contrario se requiere utilizar otros tipos de procedimientos, como los ya mencionados. Mención especial merece el caso de la confiabilidad de consistencia interna. Este índice de confiabilidad nos permite determinar hasta dónde los ítemes de una prueba son consistentes entre sí. Por lo tanto, la consistencia interna de la prueba no se refiere a la estabilidad de las decisiones que se deriban de los resultados, sino que sólo está relacionada con las características de los ítemes. En consecuencia, para las PRC la consistencia interna puede ser vista como el vehículo a través del cual se puede verificar la congruencia ítem-dominio de un conjunto de reactivos. Por lo tanto, el enfoque de consistencia interna, como se usa en las PBN, no es particularmente útil cuando se piensa en este mismo concepto, en el contexto de las PRC.

**Validez.** En las PRC típicamente se utilizan tres tipos de validez como son: la validez descriptiva, la validez funcional y la validez selección-

dominio. Mediante la validez descriptiva se intenta verificar hasta dónde una PRC está realmente midiendo lo que su esquema descriptivo sugiere. Este tipo de validez es altamente relevante para las PRC, por cuanto la misma se refiere a un aspecto sustantivo de este tipo de pruebas, como lo es la presentación de descripciones claras de los dominios de conducta que pretende medir el instrumento. Ello ofrece una mejor manera de interpretar el desempeño del sujeto en la prueba, en comparación el uso equivalente que nos permiten las PBN.

Mediante las especificaciones de la prueba se intenta establecer las reglas que servirán de referencia para: (a) la construcción de los ítemes del instrumento, a objeto de que se pueda lograr la debida congruencia item-dominio; (b) la interpretación del significado del desempeño de los sujetos en la prueba. Si las especificaciones de la prueba cumplen a cabalidad con estos dos propósitos, se puede asumir razonablemente que la misma tiene validez descriptiva.

La **validez descriptiva** de una PRC se establece en dos fases. Así, primero, se determina hasta dónde la prueba cumple con el propósito de servir de una fuente inequívoca de referencia para los diseñadores de la prueba. Para tal fin, se pueden utilizar los siguientes dos procedimientos:

El primer procedimiento consiste en: (a) seleccionar un grupo de expertos en el tema y pedirle que formulen uno o más ítemes de acuerdo con las especificaciones de la prueba; (b) determinar la homogeneidad de los reactivos producidos por los expertos. Para ello, se proceden dos maneras:

1. Se ordenan aleatoriamente los ítemes producidos por los expertos y se le pide a tres jueces independientes que evalúen la homogeneidad de los reactivos, por simple inspección. Supongamos que inicialmente fueron elaborados 10 ítemes y que al ser evaluados por los tres jueces se

encuentran los siguientes índices de homogeneidad: juez 1, 0,90 (9 de 10); juez 2, 0,80 (8 de 10); y juez 3, 1,00 (10 de 10), para un índice promedio de homogeneidad de 0,90. Este resultado puede ser interpretado como un índice aceptable de validez descriptiva de la PRC.

2. El segundo procedimiento consiste en administrar los 10 ítemes a un grupo representativo de alumnos (sujetos que ya hayan recibido el programa de instrucción que se pretende evaluar). Luego, se analizan los ítemes y se determina su índice de homogeneidad (método de consistencia interna).

En la segunda fase de validación se intenta establecer hasta dónde los ítemes de la PRC son congruentes con el dominio de conducta descrito. En tal sentido, se selecciona un grupo de jueces expertos y se les pide evaluar la congruencia de los reactivos y luego se calcula un índice de congruencia interjueces. Por ejemplo, supongamos que disponemos de los 10 ítemes referidos anteriormente y que hemos seleccionado dos jueces para el propósito en cuestión, cuyos resultados se presentan en el Cuadro 30.

*Cuadro 30. Clasificación de dos distribuciones de datos con base en la mediana*

Item	Juez 1	Juez 2
1	Si	Si
2	Si	Si
3	No	Si
4	No	No
5	Si	Si
6	Si	Si
7	Si	Si
8	Si	Si
9	Si	Si
10	Si	Si

Como se puede observar en el Cuadro 30 , el juez 1 encontró que eran congruentes 8 de 10 ítemes (80 por ciento) de la prueba; mientras que el juez 2, identificó 9 de 10 (90 por ciento). En promedio, se evidenció un índice de congruencia interjueces de 85 por ciento, el cual puede ser considerado como aceptable a los fines de una PRC.

En segundo lugar, tenemos la **validez funcional**, mediante la cual se intenta determinar hasta dónde una PRC cumple realmente con la función para la cual fue creada. La validez equivalente en las PBN es la validez referida a un criterio, en la que se intenta predecir el desempeño de un sujeto en una determinada situación futura, a partir de la ejecución actual en la prueba objeto de validación. Sin embargo, es importante señalar que la validez funcional no se refiere sólo a la validez predictiva, sino que va mucho

más allá. En este sentido, habrán tanto criterios para evaluar la validez funcional de las PRC, como propósitos se persigan con dichas pruebas.

Una PRC puede tener una excelente validez descriptiva y no tener validez funcional. Por ejemplo, podríamos descubrir, para nuestro desaliento, que los egresados quienes han tenido un buen desempeño en las pruebas administradas inmediatamente después de la instrucción, tienen una ejecución inadecuada cuando actúan en la vida real. En este caso, la prueba está mostrando que pareciera no tener validez funcional.

La diferencia entre validez descriptiva y funcional es particularmente importante para los evaluadores educacionales, ya que existen situaciones en que se requiere poner el acento en la validez funcional de la prueba, pero sin sacrificar la validez descriptiva de la misma. En consecuencia, se debe mantener un buen equilibrio entre estos dos tipos de validez. Es insostenible querer aumentar la validez funcional de una prueba a expensas de sacrificar la validez descriptiva.

En tercer lugar, tenemos la **validez selección-dominio**, la cual permite verificar el nivel de eficiencia con el que el dominio de conducta (delimitado por las especificaciones de la prueba) ha sido seleccionado. En consecuencia, el aspecto primario implícito en este tipo de validez tiene que ver con el hecho de determinar hasta dónde hemos sido capaces de seleccionar, entre varios dominios alternativos disponibles, aquél que puede servir como un indicador seguro de otras dimensiones más generales, bajo consideración. Es decir, que la posibilidad de generalización constituye un aspecto sustantivo del tipo de validez selección-dominio.

El tipo de validez equivalente en las PBN es la validez constructiva. Sin embargo, en el caso de las PRC, el tipo de validez selección-dominio no está interesada en establecer la existencia de un determinado constructo

hipotético. Existen varios procedimientos mediante los cuales se puede estimar la validez selección-dominio de una PRC. Uno de ellos puede ser expresado en los términos siguientes:

1. Se delimitan los posibles dominios de interés y se elaboran las pruebas correspondientes.
2. Se selecciona una muestra de sujetos y se les administran las pruebas correspondientes.
3. Se procesan los datos y se toma la decisión acerca del mejor dominio a ser seleccionado. Ilustremos con un ejemplo.

Supongamos que estamos interesados en evaluar un determinado conjunto de conductas, el cual ha sido delimitado en cuatro dominios específicos, que han sido identificados con las letras A, B, C y D; y que los resultados de la administración de las pruebas correspondientes se presentan en el Cuadro 31.

*Cuadro 31. Desempeño de una muestra de sujetos en cuatro PRC*

Alto Desempeño en los Dominios	Desempeño en otros Dominios			
	A	B	C	D
A	**	65%	82%	90%
B	82%	**	95%	100%
C	92%	61%	**	94%
D	42%	48%	61%	**

Si estuviéramos aprendiendo de la situación B, podríamos apreciar que quienes tienen un desempeño alto en dicha dimensión también lo tienen en el resto de los dominios, lo cual es un indicador de su potencial de generalización. En este sentido, esta dimensión presenta la mayor validez

selección-dominio; mientras que la peor está representada por el dominio identificado con la letra D.

Otra manera de estimar la validez selección-dominio consiste en: (a) diseñar un pequeño experimento en el que se enseñe deliberadamente a un grupo de alumnos a objeto de que logren la maestría correspondiente en los diferentes dominios de interés; (b) medir el poder de generalización del entrenamiento alcanzado hacia otros dominios; y (c) observar cuál de los diferentes dominios discrimina mejor entre los sujetos entrenados versus los no entrenados. En el caso de objetivos afectivos, se podría chequear el poder de discriminación de los diferentes dominios en relación con esos sujetos que han sido juzgados como positivo y negativo con respecto al dominio afectivo que se está midiendo.

# Construcción de Cuestionarios

La palabra **cuestionario** proviene del latín **quaestionarius**, que significa “lista de cuestiones que se proponen con cualquier fin”. También puede ser definida como: “lista de preguntas relativas a un mismo tema destinadas a ser contestadas por un gran número de individuos” (Salvat, 1968, p. 224). Berdie y Anderson (1974) han definido el cuestionario como “un procedimiento que permite obtener respuestas a preguntas, mediante el uso de un formulario que el sujeto puede llenar por si solo” (p. 11). Asimismo, Hernández, Fernández y Baptista (1991) han definido el cuestionario como “un conjunto de preguntas respecto a una o más variables a medir” (p. 285).

De las definiciones anteriores se concluye que un cuestionario es un instrumento de recolección de datos, de lápiz y papel, integrado por preguntas que solicitan información referida a un problema, objeto o tema de investigación, el cual es normalmente administrado a un grupo de personas. Este tipo de instrumento es de uso generalizado en los estudios de naturaleza económica, política, sociológica y psicológica.

El propósito del cuestionario y el tipo de información que solicita varía con cada caso de investigación. Por ejemplo, el Ministerio de Fomento y el Banco Central de Venezuela, emplean el cuestionario para medir la variación trimestral de los índices de precios al consumidor. Asimismo, la Oficina Central de Estadística e Información (OCEI), de la Presidencia de la República, utiliza el cuestionario para obtener información censal, cada diez

años, sobre una serie de variables demográficas, educacionales, económicas y de vivienda, de la población venezolana.

Las organizaciones políticas utilizan el cuestionario para estimar la intención del voto de la población en relación con los candidatos que compiten por determinados cargos públicos de elección popular. Las organizaciones económicas emplean el cuestionario con diferentes propósitos, por ejemplo, realizar estudios de mercado para lanzar un nuevo producto de consumo masivo; evaluar la satisfacción del público en relación con el consumo de un producto o con respecto a la prestación de un determinado servicio.

El uso del cuestionario en la investigación social se basa en un supuesto básico, cual es que el individuo dará respuestas verdaderas; es decir, se asume que el sujeto tiene tanto la voluntad como la capacidad para emitir respuestas sinceras y verdaderas acerca del objeto de estudio. Este supuesto es de vital importancia en la investigación basada en cuestionarios. Por ejemplo, algunos ítemes de un cuestionario pueden requerir del sujeto buscar información de archivo para emitir la respuesta esperada. Ahora bien, ¿Está la gente dispuesta a tomarse el tiempo necesario para hacer esto? De no ser así, sólo se deberían formular preguntas basadas en el conocimiento directo que tiene el respondiente.

Sin embargo, aun en este caso existe otro problema como es el de la posible falla de la memoria (problemas de recuerdo asociados con el fenómeno del olvido). Al respecto, algunos autores consideran que el recuerdo puede diferir de los hechos y, por lo tanto, no debería ser tomado como tales (Payne, 1951, citado por Berdie y Anderson, 1974). Incluso, aun en los casos en que la pregunta se refiere a datos numéricos, las respuestas tienden a ser incorrecta.

Tal situación puede deberse a preguntas formuladas vagamente o a la utilización de términos que no han sido previamente definidos. El problema de los datos factuales erróneos puede evitarse mediante la obtención de la información directamente de los archivos o de otras fuentes directas de información. Por esta razón, los cuestionarios deberían ser sólo utilizados en aquellos casos en que la información de interés no puede ser obtenida directamente de otras fuentes.

A menudo se encuentra en la literatura que los resultados de estudios basados en cuestionarios tienden a diferir marcadamente unos de otros. Ello podría deberse al hecho de que algunas veces se realizan comparaciones de resultados que han sido obtenidos través de cuestionarios diseñados con propósitos diferentes, con distintas personas, en tiempos diferentes. De lo anterior se deduce que cada estudio basado en cuestionario es único y debe ser concebido para llenar una necesidad específica en el marco de un estudio en particular.

En general, los cuestionarios se utilizan con dos propósitos básicos, como son: la descripción y la medición. La información adquirida a través de este tipo de instrumento puede suministrar información descriptiva de las características de individuos o grupos en relación con variables tales como: sexo, edad, años de educación, ocupación, ingresos, estado civil, vivienda, afiliación política y preferencia religiosa, entre otras.

La función de descripción puede cumplir varios propósitos, por ejemplo, el conocimiento de la distribución de la variable edad en un determinado grupo de trabajadores puede permitir al investigador dar una interpretación plausible a ciertos problemas relacionados con el desempeño laboral. También, el conocimiento de las características educacionales del grupo puede ayudar a explicar el por qué de determinadas actitudes manifestadas

por sus miembros. En síntesis, la descripción precisa de los elementos de cualquier situación social puede beneficiar al investigador de muchas maneras; por ejemplo, permite comprender, explicar y predecir los hechos, eventos o fenómenos.

Otra función importante que cumplen los cuestionarios es el de la medición de variables individuales o grupales, particularmente las actitudes. Los cuestionarios pueden contener uno a más reactivos formulados para medir varios fenómenos actitudinales, tales como: distancia social, percepción de la cohesión grupal, grado de prejuicio racial, permisividad sexual, religiosidad, ansiedad, clarificación de roles, alienación y otras.

### **Tipos de Cuestionarios**

Existen diferentes tipos de cuestionarios, aun cuando no se conocen normas universales que permitan evaluar la adecuación de cada uno de ellos a la situación particular en que son empleados. Determinar por adelantado la extensión, naturaleza, contenido y el método de administración de este tipo de instrumento, constituye una responsabilidad primaria del científico social. Los cuestionarios pueden ser clasificados atendiendo a diferentes criterios, sin embargo en esta obra se agruparán de acuerdo con las siguientes tres categorías: (a) el tipo de respuesta requerido; (b) el momento de la codificación y (c) la forma de administración.

**Tipo de respuesta.** De acuerdo con la primera dimensión o categoría, los cuestionarios pueden ser clasificados en: de respuestas cerradas, de respuesta semi-cerradas, de respuestas abiertas y mixtas. Los cuestionarios de respuestas cerradas son aquellos en que cada reactivo o pregunta consta de un número fijo de escogencias o alternativas de respuestas. En estos casos, se le pide al sujeto marcar la alternativa de respuesta que considere más apropiada; por ejemplo:

### *Ejemplo 17*

¿Cuál de las alternativas de respuestas que se presentan más abajo definen mejor su actitud hacia el aprendizaje de la matemática?

- a.  Disfruto mucho estudiando matemática
- b.  No me gusta asistir a las clases de matemática
- c.  Las clases de matemática definitivamente son aburridas
- d.  Le asigno un gran valor al estudio de la matemática
- e.  La matemática es la asignatura más importante del plan de estudio.
- f.  Me da lo mismo aprobar o no el curso de matemática

La utilización de este tipo de preguntas en un cuestionario depende, entre otras, de las siguientes razones:

1. Se asume que la población objeto de estudio tiene conocimiento acerca del área de contenido que abarca el cuestionario.
2. Se asume también que el investigador tiene suficiente información acerca de la muestra de estudio para ser capaz de anticipar qué clase de respuesta es más probable.

Otra manera de redactar ítemes de respuestas cerradas es cuando las alternativas a marcar se estructuran en forma de escala, por ejemplo:

### *Ejemplo 18*

¿Con qué frecuencia el director de su escuela convoca a la comunidad educativa para tomar decisiones importantes relacionadas con la vida del plantel?

- a.  Siempre
- b.  Casi siempre
- c.  Algunas veces
- d.  Rara vez
- e.  Nunca

Los **cuestionarios de respuestas semi-cerradas**, son aquellos que además de presentar diferentes alternativas de posibles respuestas, incluyen, además, una alternativa abierta identificada con la palabra: Otro(a), dejando el espacio correspondiente para responder. A continuación se presenta un ejemplo ilustrativo:

*Ejemplo 19*

Algunas de las razones para votar por el candidato del partido gobernante son:

- Honestidad en el ejercicio del poder
- Actitud democrática para gobernar
- Solidaridad con los que menos tienen
- Interés genuino para resolver los problemas
- Otra: \_\_\_\_\_

*(especifique)*

Los **cuestionarios de respuestas abiertas**, son aquellos cuyas respuestas requieren de la expresión escrita del sujeto, el cual debe elaborar claramente su respuesta. En estos casos, en lugar de anticipar posibles alternativas de respuestas, el investigador simplemente provee un espacio suficiente para la redacción de cada respuesta; por ejemplo:

*Ejemplo 20*

¿Cuál es su opinión acerca del clima organizacional de su empresa?

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Finalmente, los **cuestionarios mixtos**, son aquellos que están integrados por algunas preguntas cerradas o semi-cerradas y por otras abiertas. A continuación se presenta un ejemplo que ilustra este tipo de instrumento.

*Ejemplo 21*

<p><b>1.</b> Indique su opinión del curso, con base en los ítemes que se presentan más abajo:</p> <table><tr><td>Si</td><td>No</td><td></td></tr><tr><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td>Se lograron los objetivos previstos</td></tr><tr><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td>El contenido del curso fue apropiado</td></tr><tr><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td>El instructor mostró dominio del tema</td></tr><tr><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td>Los materiales instruccionales fueron útiles para los fines que fueron elaborados.</td></tr><tr><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td>La organización del curso fue eficiente</td></tr></table>	Si	No		<input type="checkbox"/>	<input type="checkbox"/>	Se lograron los objetivos previstos	<input type="checkbox"/>	<input type="checkbox"/>	El contenido del curso fue apropiado	<input type="checkbox"/>	<input type="checkbox"/>	El instructor mostró dominio del tema	<input type="checkbox"/>	<input type="checkbox"/>	Los materiales instruccionales fueron útiles para los fines que fueron elaborados.	<input type="checkbox"/>	<input type="checkbox"/>	La organización del curso fue eficiente
Si	No																	
<input type="checkbox"/>	<input type="checkbox"/>	Se lograron los objetivos previstos																
<input type="checkbox"/>	<input type="checkbox"/>	El contenido del curso fue apropiado																
<input type="checkbox"/>	<input type="checkbox"/>	El instructor mostró dominio del tema																
<input type="checkbox"/>	<input type="checkbox"/>	Los materiales instruccionales fueron útiles para los fines que fueron elaborados.																
<input type="checkbox"/>	<input type="checkbox"/>	La organización del curso fue eficiente																
<p><b>2.</b> Señale qué fue lo que más le gustó del curso:</p> <p>_____</p> <p>_____</p> <p>_____</p>																		
<p><b>3.</b> Indique qué fue lo que menos le gustó del curso:</p> <p>_____</p> <p>_____</p> <p>_____</p>																		
<p><b>4.</b> ¿Qué recomendaría usted para mejorar los aspectos que menos le gustaron del curso?</p> <p>_____</p> <p>_____</p> <p>_____</p>																		

**Ventajas y desventajas de las preguntas cerradas.** A continuación se incluyen algunas de las ventajas que presentan el uso de preguntas cerradas, ellas son:

1. Los ítemes de respuestas cerradas son fáciles de puntuar y codificar.
2. No se requiere que el respondiente escriba la respuesta, sino que simplemente la reconozca y la marque. En los casos en que el sujeto no

sea capaz de expresarse verbalmente de manera adecuada, este tipo de reactivos son los más indicados.

3. Este tipo de ítemes facilita la completación de todo el cuestionario. Los cuestionarios extensos con reactivos cerrados se completan más rápidamente que aquellos que requieren información escrita equivalente.

4. Si el cuestionario es enviado por correo, existe una alta probabilidad de que el sujeto pueda responder y regresar el instrumento con más frecuencia si no se requiere que él tenga que escribir las respuestas.

En cuanto a las **desventajas** de los cuestionarios con reactivos cerrados, podemos enumerar las siguientes:

1. Inhabilidad potencial del investigador para proveer al sujeto con todas las alternativas relevantes de respuestas. Este tipo de ítemes, como ya se mencionó antes, requiere de una cierta familiaridad con la población bajo estudio. Si el respondiente es forzado a seleccionar una alternativa, entre varias, que no se adapta a su preferencia, el investigador puede terminar con una información confusa, poco clara o irrelevante.

2. Este tipo de reactivos especialmente cuando se utilizan para medir actitudes, pueden inducir al sujeto a responder en forma tendenciosa, marcando hacia los valores extremos o centrales de las alternativas de respuesta.

**Ventajas y desventajas de las preguntas abiertas.** Las **ventajas** más resaltantes de los cuestionarios de reactivos abiertos son:

1. Son muy útiles en aquellos casos en que el investigador tiene poca información acerca de la población objeto de estudio. En este caso, el

respondiente está menos restringido para contestar una pregunta determinada.

2. Bajo ciertas circunstancias, las preguntas abiertas son útiles al investigador, por cuanto le permiten comprender mejor la conducta de un grupo bajo estudio. La flexibilidad de este tipo de reactivo elicitará respuestas imprevistas que permitirán al investigador aumentar la comprensión de lo que está estudiando.

En cuanto a las desventajas de los ítemes abiertos tenemos:

1. Este tipo de reactivos son más difíciles para la clasificación y codificación de las respuestas. Diferentes respondientes pueden emitir respuestas aparentemente similares en un ítem de un cuestionario, pero puede haber importantes diferencias en cuanto a la importancia y al significado que el sujeto le asigna a las respuestas.

2. En los reactivos abiertos existe la posibilidad de sesgo, debido a diferentes razones; en primer lugar, porque existen diferencias educacionales en los sujetos de la muestra, las cuales se expresan en limitaciones para expresarse apropiadamente tanto por escrito como oralmente. En segundo lugar, las diferencias socioeconómicas y culturales también pueden inducir sesgo en las respuestas de los ítemes abiertos, por cuanto ello hace que los sujetos tengan percepciones distintas del objeto de estudio.

3. Este tipo de reactivos consumen bastante tiempo para ser respondidos y para ser puntuados.

### **Momento de la Codificación**

De acuerdo con la manera como los reactivos son estructurados, en relación con la codificación de las respuestas para su posterior análisis, los cuestionarios pueden ser clasificados en dos grupos: pre-codificados y post-

codificados. El primer caso, se refiere a aquel tipo de instrumento cuyas preguntas están formuladas de tal manera que las alternativas de respuestas han sido previamente identificadas a través de un código numérico, previo a su administración. A continuación se presentan los casos de las variables: sexo, estado civil y nivel de instrucción, las cuales han sido previamente codificadas. El código, en cada una de dichas variables está representado por el valor numérico que antecede la alternativa de respuesta, como se aprecia en el Cuadro 32.

*Cuadro 32. Codificación de Variables*

Código	Sexo	Código	Estado Civil	Código	Nivel de Instrucción
1	<input type="checkbox"/> varones	1	<input type="checkbox"/> soltero	1	<input type="checkbox"/> Preescolar
2	<input type="checkbox"/> hembras	2	<input type="checkbox"/> casado	2	<input type="checkbox"/> Básica
		3	<input type="checkbox"/> viudo	3	<input type="checkbox"/> Media
		4	<input type="checkbox"/> divorciado	4	<input type="checkbox"/> Superior

El procedimiento anterior es particularmente útil a los fines del procesamiento electrónico de los datos, ya que las alternativas de respuestas deben ser representadas en el computador a través de códigos numéricos. En tal sentido, cuando vamos a registrar la respuesta dada, por ejemplo, a la variable sexo, transcribiremos en el computador los valores 1 ó 2, según que el sujeto haya respondido marcando en el paréntesis la alternativa varón (1) o hembra (2). Al respecto, es importante señalar que tales números utilizados como códigos en las variables, no tienen un valor de cantidad, ni de orden, sino meramente de representación de las alternativas de respuestas; es decir, su valor es nominal o clasificatorio.

Se considera que un cuestionario es postcodificado cuando la

codificación de las variables se realiza después que el instrumento ha sido administrado y respondido por el grupo de informantes. Normalmente, esta actividad se realiza previo a la transcripción de los datos al computador para su análisis.

### **Forma de Administración**

Existen básicamente cuatro formas de administrar los cuestionarios; ellas son: autoadministrado, por entrevista personal, por entrevista telefónica y, autoadministrado y enviado por correo postal, electrónico o servicio de mensajería. El cuestionario **autoadministrado** consiste simplemente en suministrar el instrumento directamente a los sujetos para que respondan, de acuerdo con las instrucciones indicadas en el mismo. En este caso no hay intermediarios, sino que ellos mismo marcan o elaboran las respuestas. Esta modalidad de administración del cuestionario presenta limitaciones cuando la muestra del estudio está formada por sujetos analfabetas o por niños pequeños.

En los cuestionarios administrados a través de la modalidad de **entrevista personal**, se requiere que un entrevistador aplique el instrumento a los respondientes. Para tal fin, el entrevistador formula las preguntas al respondiente y registra las respuesta de éste en el formulario correspondiente. Para el uso de esta modalidad de aplicación del cuestionario se requiere de un equipo de entrevistadores bien entrenados, quienes deberán estar capacitados en la técnica de la entrevista y tener un buen conocimiento del instrumento.

En cuanto a la administración del cuestionario a través de la **entrevista telefónica**, el procedimiento es similar al anterior, con la diferencia de que en este caso la entrevista no es cara a cara. Finalmente, para el procedimiento de **autoadministracion y envio postal**, el cuestionario es enviado por correo a los respondientes, éstos lo responden (autoadministración) y lo regresan por

la misma vía. En este caso, no hay retroalimentación inmediata; por lo tanto, si los sujetos tienen alguna duda en cuanto a la comprensión de las preguntas, no existe la posibilidad de que se le aclare en el momento, lo cual puede influir en la calidad de la respuesta. Las instrucciones para completar el cuestionario son generalmente enviadas en sobre cerrado. El investigador espera un tiempo prudencial (una o dos semanas) hasta que el cuestionario sea regresado. Aunque la tasa de respuesta de los cuestionarios enviados por correo puede variar, algunos autores consideran razonable esperar una tasa de retorno del 75 por ciento de los cuestionarios debidamente respondidos.

Esta modalidad de investigación mediante el uso del cuestionario enviado por correo es bastante popular en los Estados Unidos de Norte América, especialmente en la investigación de tipo comercial. Sin embargo, ha tenido poco desarrollo en Venezuela, entre otras cosas porque el servicio de correo no es lo suficientemente confiable. La modalidad **cara a cara**, llamados también de respuesta indirecta, requiere que el sujeto informante responda el instrumento en presencia del investigador o del personal asistente; es decir, esta modalidad del cuestionario requiere de un entrevistador, como ya se indicó anteriormente.

## **Pasos para el Diseño de un Cuestionario**

Un cuestionario es algo más que una lista de preguntas ordenadas en un trozo de papel, con el propósito de obtener información de un grupo de informantes acerca de un problema o tema, como pudiera parecer a cualquier persona. Para el investigador social, el cuestionario es un instrumento de recolección de datos, cuya elaboración implica una tarea compleja, ardua y, a veces, tediosa. Para diseñar un buen cuestionario se requiere de conocimiento y experiencia en el campo de las ciencias sociales y, especialmente sobre los métodos de investigación y la teoría de medición. Las fases principales para su elaboración son: (a) reflexiones previas; (b) planificación; (c) elaboración del cuestionario; (d) estudio piloto; (e) administración; y (f) análisis de los datos.

### **Reflexiones Previas**

Antes de iniciar el proceso laborioso de diseñar un cuestionario es necesario hacer algunas reflexiones, las cuales pueden estar asociadas con preguntas tales como las siguientes:

1. ¿Cuál es el propósito del cuestionario?
2. ¿A qué población está dirigido? ¿Cuáles son sus características?
3. ¿Cuan accesible es la población objeto de estudio?
4. ¿Qué modalidad de cuestionario será utilizada?
5. ¿Qué clase de reactivos será empleada?
6. ¿Qué extensión se anticipa que podrá tener el cuestionario?
7. ¿Cuánto control puede ser ejercido sobre los factores que pueden afectar la validez de los resultados?

Es importante que la finalidad que persigue el investigador con la

aplicación de un cuestionario sea hecha explícita, ya que ello permite centrarse en el aspecto sustantivo del estudio, al delimitar suficientemente su objeto y, de esa manera, evitar incluir información irrelevante o innecesaria o excluir preguntas claves o importantes. Por ejemplo, se podría diseñar un cuestionario cuyo propósito sea *determinar el estrato socioeconómico de la población de la zona metropolitana de la ciudad de Barquisimeto*.

Idealmente, la muestra de sujetos utilizados en un estudio mediante cuestionario, debería ser obtenida a través de procedimientos de muestreo estadístico, lo cual permite: (a) asegurar que cada uno de los elementos de la población tenga igual probabilidad de ser seleccionado, lo cual, a su vez, garantiza niveles aceptables de representatividad; y (b) garantizar la utilización de un tamaño adecuado de la muestra para un nivel de precisión dado, lo cual permite mantener los costos bajo control.

Tener conocimiento previo sobre las características de la población (estrato socioeconómico, educación, edad) es importante porque permite redactar los ítems a un nivel de comprensión y modalidad (verbal, gráfica, pictórica, figural) apropiados. La accesibilidad de la población también es un dato clave porque ayuda a definir la modalidad de administración del cuestionario.

El contenido del cuestionario estará determinado por la clase de información que deseamos obtener de la población bajo estudio. Es común utilizar en los cuestionarios una sección de datos de identificación que está relacionada con variables tales como: Nombre (opcional), cédula de identidad, edad, sexo, educación, ocupación, entre otras. La modalidad de administración influye en la manera de diseñar el cuestionario, particularmente en la selección del tipo de ítem y en su extensión.

La extensión del cuestionario es un asunto controversial en la literatura

especializada, que no ha logrado una solución definitiva. Algunos investigadores están a favor de los cuestionarios cortos, bajo el argumento de que entre menos extensos sean, hay más probabilidad de que aumente la tasa de respuesta. Esto es particularmente válido en el caso de los cuestionarios postales. Sin embargo, algunos autores piensan lo opuesto y están a favor de los cuestionarios largos (ver, por ejemplo, Champion y Sear, 1969, citado por Black y Champion, 1976). Estas dos posiciones parecen extremas y tienen poca utilidad práctica. Hablar de cuestionario corto o largo es una noción imprecisa y subjetiva que no ayuda a la toma de decisiones en este respecto. La recomendación que damos en esta obra es que los cuestionarios deben tener la extensión que sea necesaria en función de su propósito, la modalidad de administración, y el tiempo disponible para responder de los sujetos informantes.

El control que se pueda ejercer sobre los factores que condicionan la obtención de resultados confiables y válidos es algo que nunca debemos desestimar en un estudio mediante cuestionario. De allí que además de los factores intrínsecos relacionados con la construcción misma del instrumento sea necesario prestar atención a aspectos externos tales como: la confianza del informante, apoyada en la garantía del secreto estadístico, la disponibilidad de tiempo, la motivación del sujeto para responder y el ambiente apropiado para la administración, entre otras.

### **Planificación del Cuestionario**

La fase de planificación supone tomar en cuenta los aspectos siguientes: objetivos específicos o dimensiones del estudio, operacionalización de las variables y el diseño del cuestionario. Los objetivos específicos se refieren a las dimensiones o áreas particulares de información que se requiere, de acuerdo con lo establecido en el propósito del instrumento. Por ejemplo, en el

caso anterior, en el cual el propósito era determinar el estrato socioeconómico de la población de la zona metropolitana de la ciudad de Barquisimeto, los objetivos específicos están relacionados con las cuatro áreas o dimensiones que sirven de base para estimar el estrato socioeconómico, como son: la ocupación, el ingreso, la educación y condiciones de alojamiento, de acuerdo con el método Graffar, modificado por Mendez-Castellanos (1985).

Una vez identificadas las dimensiones en que se divide el propósito del cuestionario, las variables de estudio se operacionalizan a través de sus indicadores. En el ejemplo anterior, las variables pueden ser operacionalizadas de la siguiente manera: la **ocupación**, mediante los siguientes indicadores:

- Obrero no calificado
- Obrero calificado
- Técnico medio
- Técnico superior
- Profesional universitario

De la misma manera la variable **ingreso**, podría ser operacionalizada con base en la cantidad de dinero que recibe cada persona periódicamente (semanal, quincenal, mensual, anual), como pago por la prestación de un servicio en el sector público o privado. El ingreso también podría ser representado a través de una escala que exprese, en forma de categoría, los diferentes niveles de ingreso que tiene la población, como se expresa a continuación:

- 144.000 o menos
- 144.001 a 288.000
- 288.001 a 432.000
- 432.001 a 576.000
- 576.001 y más

La variable **educación**, puede ser operacionalizada de varias maneras, dependiendo del nivel de precisión deseado en la medición de la variable. Así tenemos que dicha variable se puede expresar con valores continuos tomando en cuenta el número de años de estudios aprobados (por ejemplo, 3, 6, 9 y así sucesivamente). También se puede expresar como una variable discreta identificando la condición de alfabeto versus analfabeto. Dicha variable puede ser expresada, igualmente, como una función de los niveles de estudios aprobados, como se indica a continuación:

- Educación superior completa
- Educación superior incompleta
- Educación media completa
- Educación media incompleta
- Educación básica completa
- Educación básica incompleta

La variable **condiciones de alojamiento**, puede ser operacionalizada a través de los siguientes indicadores:

- Vivienda con óptimas condiciones sanitarias en ambientes de lujo.
- Vivienda con óptimas condiciones sanitaria en ambiente sin exceso de lujo, espacioso.
- Vivienda con buenas condiciones sanitarias en espacios reducidos.
- Vivienda con ambientes espaciosos o reducidos y con deficiencias en las condiciones sanitarias.
- Rancho o vivienda con espacio insuficiente y condiciones sanitarias marcadamente inadecuadas.

Finalmente, una vez operacionalizadas las dimensiones o variables del cuestionario, se pasa a formular las preguntas o reactivos del mismo, como se indica en el Cuadro 33.

*Cuadro 33. Plan de operacionalización de las variables*

Propósito	Dimensión	Indicadores	Ítemes
Determinar el estrato socioeconómico de la población de la zona metropolitana de la ciudad de Barquisimeto	Ocupación	<ol style="list-style-type: none"> <li>1. Profesional Universit.</li> <li>2. Técn. Sup.</li> <li>3. Técn. Med.</li> <li>4. Obrero Esp.</li> <li>5. Obrero No Esp.</li> </ol>	¿Cuál es la ocupación del jefe de familia?
	Educación	<ol style="list-style-type: none"> <li>1. Universit.</li> <li>2. Med. Completa</li> <li>3. Med. Incompleta</li> <li>4. Primaria</li> <li>5. Analfabeta</li> </ol>	¿Cuál es el nivel educativo del jefe de familia?
	Ingreso	<ol style="list-style-type: none"> <li>1. Fortuna heredada</li> <li>2. Honorarios</li> <li>3. Sueldo mensual</li> <li>4. Salario semanal</li> <li>5. Donaciones Públicas</li> </ol>	¿Cuál es el ingreso familiar?
	Condiciones de Alojamiento	<ol style="list-style-type: none"> <li>1. Vivienda con ECS, lujosa</li> <li>2. Vivienda con ECS, sin lujo</li> <li>3. Vivienda con BCS, espacios reducidos</li> <li>4. Vivienda con DCS</li> <li>5. Rancho o barraca</li> </ol>	¿Cuál es a condición de alojamiento del grupo familiar?

### Elaboración del Cuestionario

Al momento de elaborar el cuestionario es conveniente que el investigador reflexione sobre aspectos tales como los siguientes: tipo de preguntas a ser utilizadas (cerradas, abiertas, mixtas), cuántas preguntas utilizar por dimensión o variable, si habrá o no precodificación de las variables. También

se deben tomar en cuenta las características que presentan las preguntas de un buen cuestionario. Al respecto, Hernández y cols. (1991) han sugerido las siguientes:

1. **Las preguntas deben ser claras y comprensibles.** En este sentido, se debe evitar utilizar términos confusos o ambiguos en la redacción de los reactivos; por ejemplo, la pregunta ¿asiste usted a la Iglesia? quedaría mejor redactada si se preguntara: ¿con que frecuencia va usted a su Iglesia? (diariamente, los fines de semana, ocasionalmente, durante las celebraciones especiales, nunca). Posteriormente, se podría preguntar ¿a qué Iglesia asiste?, ¿dónde? ¿en qué horario?

2. **Las preguntas no deben incomodar al respondiente.** Las preguntas acerca de temas tales como la edad, el homosexualismo, la drogadicción, el uso de anticonceptivos, la prostitución y la pornografía, entre otros, por lo general tienden a incomodar al sujeto cuando se le formulan, lo cual podría llevar a que el individuo evite la respuesta o que responda de acuerdo con su mejor conveniencia, en lugar de emitir una respuesta sincera. En consecuencia, es necesario indagar sobre estos temas de manera sutil e indirecta. Por ejemplo, en el caso de la edad, especialmente, cuando se pregunta a las damas, en lugar de formular la pregunta: ¿Cuántos años tiene usted?, se podría preguntar, ¿Cuál es su fecha de nacimiento? y deducir la edad a partir de esa información. Otra manera sería preguntar: ¿En cuál de las siguientes categorías se encuentra su edad:

- Menos de 20 años
- Entre 21 y 30

- Entre 31 y 40
- Entre 41 y 50
- Entre 51 y 60
- Más de 60 años

3. **Las preguntas deben referirse a un sólo aspecto o asunto.** Formular una pregunta como la siguiente: ¿Con qué frecuencia va usted a la playa y a la montaña? (frecuentemente, rara vez, nunca), puede resultar confuso para el respondiente, ya que la frecuencia con que va a la playa puede ser distinta a la de la montaña. Por tanto, en estos casos sería preferible hacer dos preguntas, en lugar de una.

4. **Las preguntas no deben inducir las respuestas.** Cuando se diseña un cuestionario se deben evitar aquellos reactivos tendenciosos, que sugieren un tipo de respuesta. Por ejemplo, ¿Considera usted a William Manzano mejor candidato que Carlos Sifontes para dirigir nuestra asociación de investigadores? Esta pregunta es sesgada; sería más conveniente preguntar ¿Cuál de los siguientes será el mejor candidatos para dirigir nuestra asociación de investigadores:

- William Manzano
- Carlos Sifontes

5. Las preguntas no deben apoyarse en instituciones, ideas respaldadas socialmente ni en evidencia comprobada. Al respecto, Hernández, Fernández y Baptista (1993) han propuesto el ejemplo

siguiente: la Organización Mundial de la Salud ha realizado diversos estudios y concluyó que el tabaquismo provoca diversos daños al organismo, ¿usted considera que fumar es nocivo para su salud? Esta forma de pregunta sesga la respuesta, por eso no se debe emplear en el diseño de los cuestionarios.

**6. El orden fijo de las alternativas de respuesta puede favorecer determinadas opciones.** Por ejemplo, en el caso de los instrumentos en los que los valores escalares están siempre colocados en el mismo orden, como en el caso en el que se utiliza la escala siguiente:

- Siempre
- Casi siempre
- Algunas veces
- Rara vez
- Nunca

En este caso el sujeto respondiente puede tender a marcar el valor central (3) o los extremos (5 ó 1) de la escala.

**7. El lenguaje debe ser apropiado al nivel de los respondientes.** En este sentido, se debe tomar en cuenta variables tales como la edad, el nivel educativo, el estrato socioeconómico, entre otras, ya que las mismas harán variar la modalidad y estilo en la formulación de las preguntas. Por lo tanto, si hemos seleccionado una muestra heterogénea en cuanto a las variables ya señaladas, sería conveniente utilizar un lenguaje que pueda ser accesible a todos los sujetos integrantes de la muestra.

**El formato del cuestionario.** En el proceso de elaboración del cuestionario hay una fase que se refiere al arte de diseñar el instrumento, la cual incluye dos aspectos; en primer lugar, lo diagramático; es decir, la manera como se presentan y distribuyen las preguntas en el papel. Es importante que las preguntas estén organizadas de tal manera que resulten estéticamente bien presentadas, y lógicamente ordenadas. Todo ello influirá para que el cuestionario sea interesante y fácil de aplicar. Al respecto, Aaker y Day (1989) consideran que en el caso del cuestionario autoadministrado la calidad del papel, la claridad de la reproducción y la apariencia de amontonamiento son variables que deben ser tomadas en cuenta.

En segundo lugar, se debe considerar la secuencia de las preguntas. Investigaciones al respecto han sugerido que el orden de las preguntas está relacionado con la confiabilidad de la respuesta que emiten los sujetos (ver Cantril, citado por Ander-Egg, 1980). Algunos autores consideran que el orden de las preguntas en el cuestionario está determinado inicialmente por la necesidad de obtener y mantener la cooperación del entrevistado así como por hacer el cuestionario lo más fácil posible para que lo aplique el entrevistador. Luego, se debe prestar atención al problema del **sesgo de ordenamiento**; es decir, la posibilidad de que las preguntas anteriores influyan en las respuestas de las preguntas subsecuentes (ver Black y Champiom, 1976, Berdie y Anderson, 1974).

Aaker y Day (1989) sugieren que para lograr que un cuestionario sea interesante y lógico, se sigan las recomendaciones siguientes:

1. Se debe empezar con una pregunta fácil y no agresiva. La idea es ganarse la confianza del entrevistado para que coopere. Por ejemplo, se podría iniciar así: “Estamos haciendo una encuesta sobre las necesidades de capacitación del personal docente de Educación Básica en el Estado

Lara, la primera pregunta es: ¿En que área del conocimiento le gustaría recibir algún curso de capacitación, actualización o mejoramiento?

2. El cuestionario debe fluir sutil y lógicamente de un tema al siguiente. Los cambios repentinos en los temas deben evitarse, pues tienden a confundir a los entrevistados y a causar indecisión. Cuando se introduce un nuevo tema debe hacerse una afirmación de cambio o pregunta de transición, explicando la forma como el nuevo tema se relaciona con el tema anterior o con el propósito del estudio.

3. En la mayoría de los casos es mejor avanzar de preguntas amplias y generales a las más específicas. En tal sentido, se podría empezar preguntando ¿Qué piensa usted de la educación en el Estado Lara? ¿Cuáles cree usted son los problemas educativos más urgentes?, para luego proseguir, ¿Qué opina usted de la calidad de los docentes de Educación Básica? y por último: ¿Cree usted que sería necesaria poner en práctica un programa de capacitación docente? Esta manera de secuenciar los reactivos ayuda al entrevistado a ubicar la pregunta específica en un contexto más amplio y a dar una respuesta más reflexiva.

4. Las preguntas confidenciales o difíciles, tales como, por ejemplo, el nivel de ingreso y la habilidad, no deben ser colocadas al inicio del cuestionario. Deben ser introducidas, más bien, cuando el sujeto haya adquirido alguna confianza con el entrevistador y el estudio. En las entrevistas cortas pueden ser pospuestas hasta el final del cuestionario.

### **Estudio Piloto**

Una vez que la primera versión del cuestionario ha sido concluida, es necesario verificar su funcionalidad en un grupo de sujetos que no formarán parte de la muestra definitiva del estudio, pero que tienen características similares a aquéllos. Es lo que se denomina la **prueba** o **estudio piloto** del

cuestionario, el cual se realiza en dos fases: (a) validación por juicios de expertos; y (b) uso de una muestra piloto de sujetos. En el primer caso, el procedimiento a utilizar es similar al empleado para determinar la validez de contenido de un instrumento. En el segundo caso, por lo general, se considera que el tamaño de la muestra para estos fines puede oscilar entre 30 y 50 sujetos, seleccionados al azar.

Con la información obtenida de la muestra piloto se pueden identificar errores típicos que se cometen en el proceso de elaboración de los cuestionarios, particularmente en lo que respecta a la redacción de las preguntas; por ejemplo, en cuanto a ambigüedad, definiciones incorrectas, sesgo y de doble efecto. Este paso, es importante porque permite chequear la confiabilidad y validez de las preguntas.

Una pregunta es confiable si ella es capaz de evocar respuestas de una manera consistente, lo cual se manifiesta cuando una persona responde una pregunta de la misma manera en subsecuentes entrevistas. Como el lector puede suponer, tal consistencia en la respuesta dependerá, entre otros factores, de la interpretación que el sujeto respondiente haga de la pregunta, lo cual, a su vez, está asociado con la calidad técnica del reactivo en cuanto a su redacción.

Por ejemplo, si a través de la pregunta de un cuestionario se intenta obtener información acerca de las características relevantes que tienen tres personas para ser gerentes de un plantel escolar y la pregunta se formulara así: ¿Por cuál de los tres candidatos que se mencionan más abajo (A, B, y C) votaría usted para la dirección del plantel?

Si la pregunta anterior se hiciera a una muestra de 30 docentes, cada uno de ellos emitiría la respuesta por el candidato de su preferencia, pero no necesariamente por aquel que tiene las mejores características para el

desempeño exitoso en el cargo, ya que algunos pudieran seleccionar su candidato con base en un criterio meramente afiliativo (amistad, familiaridad, razón política, religiosa o ideológica), otros pudieran hacer su elección de acuerdo con la simpatía o popularidad de los candidatos; mientras que otros pudieran hacerlo con base en el conocimiento que tienen de cada uno de los aspirantes al cargo.

Todo lo anterior indica que la pregunta, tal como fue formulada, no permite obtener una respuesta confiable, de acuerdo con su objetivo específico, puesto que cada uno de los respondientes parte de criterios diferentes para responder. Por lo tanto, si lo que se quiere es saber cuál es el candidato que reúne las mejores características para desempeñarse como director de un plantel, lo mejor sería, en primer lugar, definir un perfil deseable del candidato a director.

En segundo lugar, suministrar suficiente información a los sujetos de la muestra acerca de la trayectoria profesional de los candidatos (curriculum vitae) y, en tercer lugar, formular la pregunta de tal manera que los sujetos de la muestra puedan evaluar cada candidato y, con base en ello, emitir su respuesta. Por ejemplo, una manera de formular la pregunta podría ser.

¿Cuáles son las cinco razones por las que usted seleccionaría al candidato de su preferencia para ejercer la dirección del plantel?

La validez de la pregunta, por su parte, se refiere a si el ítem es capaz o no de elicitar la respuesta verdadera y relevante, en relación con el objetivo específico del instrumento. Por ejemplo, supongamos que se está haciendo un estudio con el propósito de seleccionar grupos familiares para el programa de la Beca Alimentaria de un país, para lo cual algunos de los criterios de selección, para participar en este programa, son el número de niños en edad escolar que asisten a la escuela y el nivel socioeconómico de la familia. En

este caso, si la pregunta formulada fuera ¿Cuántos niños tiene usted y cuál es la edad de cada uno de ellos? No es una pregunta válida a los fines propuestos en el cuestionario, por cuanto ella no permite conocer cuántos niños asisten actualmente a la escuela, ni tampoco apreciar el nivel socioeconómico de la familia. Por lo tanto, la pregunta debería ser desglosada en varias, como se indica a continuación: ¿Tiene usted niños en edad escolar?, de ser cierto, ¿Cuántos de ellos asisten actualmente a la escuela? ¿En cuál escuela están inscritos? ¿Es esta escuela pública o privada? Como se puede observar la confiabilidad y la validez de las preguntas dependen mucho de la manera como el investigador redacta las preguntas del cuestionario.

Para que la prueba piloto sea efectiva, es necesario que el investigador se abra a la crítica y está dispuesto a corregir deficiencias; en tal sentido, un buen punto de partida es que él se coloque en el lugar del entrevistado y trate de contestar las preguntas. Durante una prueba previa del instrumento es recomendable que el entrevistador observe y esté muy pendiente de las reacciones del respondiente (incomodidad, resistencia) ante las preguntas del cuestionario. Al final del mismo, el entrevistador debe solicitar información al entrevistado acerca de cada pregunta (claridad, precisión, congruencia).

Con base en la opinión emitida por los sujetos de la muestra utilizada en el estudio piloto, el investigador puede modificar, sustituir o excluir algunas preguntas. En caso de que los cambios sean menores, el paso siguiente sería administrar el instrumento en la muestra del estudio. Pero, si los cambios introducidos al cuestionario como consecuencia del estudio piloto, fueran muchos, habría que considerar la posibilidad de rehacer el instrumento y someterlo a una nueva prueba piloto.

### **Observaciones sobre la Confiabilidad de los Cuestionarios**

A diferencias de otros instrumentos, tales como las pruebas o las escalas de

estimación, a los cuestionarios no se les calcula el coeficiente de confiabilidad de consistencia interna. Esta exigencia técnica se cumple indirectamente para cada uno de los ítemes, cuando se realiza el estudio piloto del cuestionario, con el propósito de determinar su grado de funcionalidad; en este caso, se hacen consideraciones sobre los aspectos siguientes: claridad de la pregunta en cuanto a su redacción, precisión en las instrucciones para responder, ausencia de sesgo en la formulación de los ítemes, adecuación idiomática y al nivel cultural de los respondientes, entre otros. Sin embargo, no es posible estimar la confiabilidad de consistencia interna del instrumento, ya que ésta expresa el grado de homogeneidad de los ítemes del instrumento, y el cuestionario, por definición, está integrado por un conjunto de preguntas que son heterogéneas entre sí. De allí que sea un contrasentido efectuar este tipo de cálculo. Veamos un ejemplo.

Suponga que queremos evaluar el desempeño del docente de educación media y utilizamos una escala de estimación integrada por 20 ítemes. Estos reactivos representan hipotéticamente diferentes aspectos que son representativos del constructo que deseamos evaluar (desempeño docente). Es decir, se espera que los ítemes guardan entre sí cierto grado de homogeneidad con respecto al constructo, pero eso no lo podemos saber sino hasta que hacemos la estimación del coeficiente de confiabilidad de consistencia interna. En este sentido, si el valor obtenido es alto (por ejemplo, entre 0,81 y 1,0) tenemos una evidencia empírica de que el instrumento cumple con esta exigencia psicométrica. Por otra parte, si lo que queremos es conocer, mediante un cuestionario, por ejemplo, el perfil demográfico-académico de dichos docentes seleccionaríamos reactivos tales como, género, edad, estado civil, nivel de educación, experiencia docente y curso que dicta, entre otros. Como se puede apreciar, estos ítemes son heterogéneos entre sí y no

representan un constructo como tal, sino que cada variable constituye un aspecto particular de la situación que interesa estudiar. En consecuencia, no es posible estimar un índice de homogeneidad en estos casos. Por esa razón, no se estima la confiabilidad de consistencia interna en este tipo de instrumento.

### **Administración**

Como ya se mencionó anteriormente en este mismo capítulo, existen diferentes modalidades para administrar un cuestionario (autoadministración, entrevista personal, entrevista telefónica, autoadministración y envío postal). Pero, independientemente de ello, es necesario haber resuelto previamente el problema de la selección de la muestra del estudio, lo cual implica, en primer lugar, definir su tamaño. Para tal fin, existen diferentes procedimientos estadísticos, dependiendo de: (a) las características de la población (finita versus infinita; homogénea versus heterogénea; grande versus pequeña); (b) propósito del estudio (estimación de totales, promedios o proporciones); (c) precisión en la estimación (asumir error grande versus pequeño); (d) nivel de confianza (90%, 95% o 99%).

En esta obra no se desarrollan los procedimientos para estimar el tamaño de la muestra ni los procedimientos de selección, por considerar que son temas estadísticos muy especializados que exceden el propósito de esta obra. Sin embargo, los interesados pueden consultar a: Aaker y Day (1989); Festinger (1978); Gabaldón Mejía (1973); Kish (1965); Seijas (1981) y Sierra Bravo (1987), entre otros. En todo caso, la recomendación del autor de esta obra es que cuando existan dudas al respecto, se consulte con un especialista, ya que no existen “recetas” para resolver el problema del tamaño de la muestra y el procedimiento de selección, sino que cada decisión debe tomarse en función de la naturaleza específica del estudio de que se trate. Para ello se

requiere tener un buen entrenamiento en el manejo de los métodos estadísticos.

En segundo lugar, se requiere seleccionar el método de muestro más apropiado (simple al azar o aleatorio simple, sistemático, estratificado, por conglomerado, entre otros) y el procedimiento de selección (tipo lotería, tabla de números aleatorios, programa computarizado, entre otros).

En tercer lugar, se requiere planificar la logística para el trabajo de campo, lo cual implica, reproducir suficiente número de formularios, seleccionar y entrenar los encuestadores (si fuera necesario), elaborar cronograma de aplicación y prever mecanismos de movilización y supervisión en el trabajo de campo.

### **Análisis de Datos**

Después que el cuestionario ha sido administrado y se estima que la tasa de retorno está dentro de los límites previsibles (aproximadamente un 75 por ciento del total de la muestra, como mínimo), se pasa a la etapa de análisis de los datos. La estrategia de análisis debe ser planificada previo al levantamiento del estudio y no al final. La manera de hacerlo dependerá, en primer lugar de la orientación del estudio, en el sentido de si su propósito es de carácter descriptivo o explicativo.

En el caso de los estudios descriptivos, el análisis de los datos puede ser realizado con base en el apoyo de estadísticos de: (a) tendencia central (frecuencias absolutas y relativas, medias, mediana, moda, curtosis, asimetría, entre otras), (b) variabilidad (amplitud, desviación semi-intercuartilar, desviación estándar, varianza, coeficiente de variación), (c) posición (deciles, percentiles), (d) coeficientes de correlación simple (Pearson, Spearman, biserial, punto biserial, Phi, tetracórica, Chi Cuadrado). También el análisis

descriptivo de los datos puede ser apoyado en el uso de cuadros y gráficos (barras, circular o tipo pastel, histogramas, polígono de frecuencias, ojiva).

Los estudios explicativos, por su naturaleza, se ubican en el campo de la estadística inferencial y, en consecuencia, requieren del uso del procedimiento de la prueba de hipótesis. En este caso, pueden ser aplicables técnicas estadísticas como: el análisis de regresión múltiple, la correlación canónica, y el análisis de varianza, entre otros. A continuación se presenta un ejemplo que ilustra ambos tipos de análisis. Suponga que un investigador desea realizar un estudio para responder las siguientes preguntas: ¿Cuál es el estrato socioeconómico de los estudiantes de quinto año de Educación Media del Estado Lara y ¿Qué relación tiene el estrato socioeconómico con el rendimiento académico de dichos estudiantes?

La primera pregunta es descriptiva porque intenta caracterizar la población de estudiantes de quinto año de Educación Media, de acuerdo con su estrato socioeconómico; mientras que la segunda es explicativa, ya que trata de indagar sobre la relación asociativa o causal del estrato socioeconómico (variable independiente) con el rendimiento académico (variable dependiente).

Suponga, además, que: (a) en el estudio se utilizó un cuestionario que incluye las siguientes variables: ocupación, educación, fuente de ingresos, condiciones de alojamiento y rendimiento académico, expresado en términos de aprobados reprobados (ver Anexo 1); y (b) el estudio se hizo en una muestra aleatoria de  $n = 100$  sujetos.

Para responder la primera pregunta, se siguió el siguiente procedimiento: (a) se elaboró una matriz de sujeto por ítem; (b) se sumaron los códigos de las variables y se obtuvo una distribución de totales para cada sujeto; y (c) se

agruparon los sujetos en estratos de acuerdo con el criterio sugerido por el método Graffar, como se indica en el Cuadro 34.

*Cuadro 34. Distribución Estrato Socioeconómico según las puntuaciones obtenidas en el Cuestionario Graffar*

Puntaje	Estrato	Denominación
4 - 6	I	Alto
7 - 9	II	Medio Alto
10 - 12	III	Medio Bajo
13 - 16	IV	Obrero
17 - 20	V	Marginal

Los resultados de la clasificación por estratos se presentan en el Cuadro 35. Como se puede observar, la población de estudiantes de quinto año de Educación Media del Estado Lara se encuentra distribuida de manera casi uniforme a través de los diferentes estratos. Si se comparan estos resultados con los de la distribución por estratos de la población nacional, la cual es aproximadamente así: Estrato I 5%, Estrato II 10%, Estrato III 15%, Estrato IV 30% y Estrato V 40%, se puede concluir que mientras los estratos alto y medio-alto están sobre representados en la población de estudiantes de Educación Media, los estratos obrero y marginal se encuentran subrepresentados.

*Cuadro 35. Distribución por estrato socioeconómico de una muestra de estudiantes de segundo año de educación media del estado Lara (n = 100)*

	Frec. Absoluta			Frec. Relativa		
	Aprob	Reprob	Total	Aprob	Reprob	Total
I (Alto)	15	5	20	15%	5%	20%
II (Med-Alto)	17	5	22	17%	5%	22%
III (Med-Bajo)	10	10	20	10%	10%	20%
IV (Obrero)	5	9	14	5%	9%	14%
V (Marginal)	7	17	24	7%	17%	24%
Total	54	46	100	54	46%	100%

Para responder la segunda pregunta del estudio se utilizó la prueba de Independencia del Chi-Cuadrado, de acuerdo con los datos que se presentan en el Cuadro 36. La formula utilizada se presenta a continuación:

$$x^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

Donde:

$X^2$  = valor de Chi Cuadrado

$F_o$  = frecuencia observada

$F_e$  = frecuencia esperada

*Cuadro 36. Distribución de los Sujetos por Estrato y Rendimiento Académico*

Variables	Factor I
Educación ( $X_1$ )	0,756
Ocupación ( $X_2$ )	0,622
Ingresos ( $X_3$ )	0,825
Vivienda ( $X_4$ )	0,589

El valor obtenido fue de  $X^2(4) = 16,31$  ( $p < 0,01$ ), lo cual indica que existe una relación significativa entre las variables estrato socioeconómico y rendimiento académico. Para determinar la magnitud de esta correlación se utilizó el coeficiente Cramer Phi, cuyo valor fue de 0,40, lo cual indica que las dos variables tienen una asociación positiva y moderada-baja. La fórmula utilizada fue:

$$\text{Cramer's } \Theta = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Donde:

$\varphi$  = es el valor de coeficiente Cramer

$X^2$  = valor de Chi Cuadrado obtenido

$N$  = número de observaciones

$K$  = número de columnas o filas (el valor más pequeño)

Como el lector puede observar no se destaca la información específica correspondiente a las variables: ocupación, educación, ingreso y condiciones de alojamiento, ya que el interés del estudio estaba centrado en la determinación del estrato socioeconómico y no en la consideración de cada variable por separado. No obstante, se puede hacer tal análisis descriptivo por separado como complemento de los resultados del estudio.

# Análisis Factorial

El conocimiento científico, en el contexto de la epistemología de la ciencia clásica, se entiende como el resultado de la reflexión crítica del científico en su intento por explicar los hechos o fenómenos de la realidad, asumida ésta como externa e independiente del sujeto cognoscente. Esta reflexión normalmente conduce a la teoría, la cual puede ser definida como: “un conjunto de constructos (conceptos) interrelacionados, definiciones y proposiciones que presentan un punto de vista sistemático de los fenómenos mediante la especificación de relaciones entre variables, con el propósito de explicar y predecir los fenómenos” (Kerlinge y Lee, 2002, p. 8).

Pero la teoría debe ser comprobada mediante su confrontación con la realidad, en el proceso de investigación. En dicho proceso, el investigador formula hipótesis y utiliza instrumentos que le permiten obtener datos para la comprobación de aquéllas. El investigador dispone de una serie de recursos analítico-metodológicos con el propósito de validar sus resultados, uno de ellos es el análisis de factores (AF), el cual permite reducir la multiplicidad de pruebas y medidas a una mayor simplicidad, al mismo tiempo que ayuda al científico a localizar e identificar las unidades o propiedades básicas que fundamentan las pruebas y medidas (Kerlinger y Lee, 2002.).

El AF puede ser entendido como un proceso formal de toma de decisiones con el propósito de explicar la estructura subyacente en una matriz de datos, al mismo tiempo que permite reducirla a dimensiones significativamente más pequeñas, lo cual facilita su interpretación. Tiene

múltiples aplicaciones en las ciencias de la conducta; una de ellas se presenta en el campo de la teoría de medición, particularmente en lo que respecta al desarrollo de instrumentos (escala de estimación, pruebas de habilidades, aptitudes, conocimientos y otras). Esta aplicación es enfatizada en este trabajo, por su utilidad en el proceso de refinación técnica y teórica de los instrumentos de medición (análisis de ítem, validez, confiabilidad, elaboración de índices). En este sentido, el AF es particularmente útil como método de validación de constructo.

La utilización del AF como método analítico en el proceso de construcción de instrumentos ofrece ventajas comparativas en relación con otros métodos estadísticos que se utilizan convencionalmente para el mismo propósito. No obstante, el AF ha sido poco difundido, en la comunidad académica internacional, dada su complejidad matemática, lo cual hace que, por lo general, sea ofrecido como un curso de estadística avanzada en estudios doctorales.

Actualmente, con el uso del computador, los problemas de cálculo, para la utilización del método de AF se han simplificado al máximo, pudiendo ser más accesible a un mayor número de investigadores interesados. No obstante, es necesario familiarizarse con los aspectos teórico-conceptuales del método a objeto de sacar mayor provecho de sus posibilidades analíticas y heurísticas.

En este capítulo se presenta de manera condensada los aspectos teóricos, metodológicos y prácticos más importantes del AF, a objeto de que pueda ser seguido con relativa facilidad por los interesados. En tal sentido, el trabajo ha sido organizado en cinco secciones, a saber: aspectos conceptuales y origen del AF, métodos de factorización, aplicaciones del AF, el proceso del AF y, AF y desarrollo de instrumentos de medición.

## **Aspectos Conceptuales y Origen del Análisis de Factores**

El análisis de factores (AF) es una técnica de análisis de datos, multivariable, que se caracteriza por su capacidad para simplificar o reducir una matriz de datos (matriz  $X$  o de sujetos por medidas) a dimensiones estructurales más pequeñas. Dada una matriz de datos, transformada en una matriz de intercorrelación (medidas por medidas), la técnica del AF nos permite apreciar hasta donde existe algún patrón subyacente de relaciones en los datos de tal manera que haga posible una reducción de las medidas a un conjunto más pequeño de factores o componentes que puedan ser tomados como las variables originarias o fuentes que explican las interrelaciones observadas en los datos.

El AF puede ser visto también como un método que permite descomponer la varianza implícita en la matriz de datos, en dos grandes componentes: varianza común y varianza única. La primera se forma con base en las combinaciones lineales de variables que están correlacionadas. El conjunto de la varianza común que una variable comparte con uno o más factores recibe el nombre de comunalidad ( $h^2$ ) y se obtiene sumando los valores al cuadrado de los coeficientes de correlación variable-factores. La varianza única ( $1 - h^2$ ) está conformada por la varianza que es específica de cada variable, más la varianza de error. El propósito del AF es extraer la varianza común e ignorar la varianza única.

En caso de que se desee mantener la varianza específica y dejar de lado sólo la varianza de error, el coeficiente de confiabilidad puede ser insertado en la diagonal de la matriz de correlación (matriz  $R$ ) para ser analizada factorialmente. En este caso, la proporción de la varianza explicada por la matriz factorial estaría constituida por la varianza confiable (común más específica). En la investigación actual la varianza puede ser descompuesta en

su parte que es común y la que es específica. Las intercorrelaciones de una batería de pruebas puede ser empleada en una ecuación para lograr los coeficientes de correlación múltiple que cuando se elevan al cuadrado expresan la proporción de la varianza total que es común. La diferencia entre las varianzas confiable y común es igual a la varianza específica del test.

En la concepción del AF está implícita la idea de que si se tiene un conjunto de variables correlacionadas entre sí, estas relaciones recíprocas podrían deberse a la presencia de una o más variables (factores subyacentes) relacionadas en grados diversos con aquéllas (Blalock, 1977). El propósito del AF es identificar estos factores o variables comunes, más generales que los datos mismos.

Un *factor* puede ser definido matemáticamente como cualquier combinación lineal de variables en la matriz de datos; por ejemplo, sea el factor A, integrado por las variables a, b, c... k y sus ponderaciones  $w_a$ ,  $w_b$ ,  $w_c$ ,...  $w_k$ , su expresión algebraica sería:

$$A = w_a a + w_b b + w_c c + \dots + w_k k$$

Si la ponderación de la variable *a* fuera 0,70 ( $w_a = 0,70$ ), la puntuación de cada sujeto en la medida *a* se multiplicaría por 0,70. Los pesos para cada variables pueden ser iguales o diferentes, con signos positivos o negativos indistintamente. Los diferentes métodos que se utilizan para derivar los factores se definen de acuerdo con la forma en que se utilizan los pesos para obtener las combinaciones lineales. Es importante tomar en cuenta que cada factor está constituido por un conjunto de puntuaciones reales (observables); esto supone que cada sujeto tiene una puntuación en el factor; por ejemplo, al combinar todas las puntuaciones del sujeto 1 en las medidas de *a* hasta *k* se obtiene su puntuación para el factor A. Para todos los sujetos se utilizan los mismos pesos, pero como cada uno tiene diferentes puntuaciones en las

distintas medidas, también obtendrá diferentes calificaciones en el factor (ver Nunnally, 1987).

Existen tres posibles interpretaciones, en la literatura, sobre los factores y su relación con las variables que lo integran, ellas son: (a) como *indicadores de efectos*, donde las variables observadas son consideradas como resultado de una variable subyacente latente (Bollen y Lennox, 1991); (b) como *componentes*, donde las variables son transformadas en otras variables por conveniencia; y (c) como indicadores causales, donde la variable latente es considerada como el resultado de las observables (Bollen y Lennox, 1991).

Los *indicadores de efectos* son combinaciones lineales en que las observables son los resultados (efectos, consecuencias) del factor. Es decir, las medidas observables, actúan como variables dependientes y el factor es una variable independiente. En este caso se asume que el factor está libre de error, pero no así las medidas. Nunnally y Bernstein (1995) ilustran esta situación de la siguiente manera “un sujeto que sale bien en una prueba que es una buena medida del factor fluidez verbal, lo logra como una consecuencia de tener fluidez verbal o debido a la suerte; salir bien en la prueba no causa que el individuo se vuelva verbalmente fluido. Además, el factor es más amplio en significado que cualquiera de sus observables falibles, ya que no es definido por completo por ellos, individualmente o en combinación” (p. 505). La mayoría de los estudios que se realizan en psicología y educación tienen este enfoque.

Los *componentes*, como ya se ha mencionado son combinaciones lineales de observables y por consiguiente observables por derecho propio. Suponga, por ejemplo, que las variables  $X_1$  y  $X_2$  son dos medidas, que representan el pretest y el posttest respectivamente en una investigación; dos

posibles componentes explicativos del desempeño global y del aprendizaje podría estar formado por  $X_3 = X_1 + X_2$ , y por  $X_4 = X_1 - X_2$ . Existen muchas otras posibilidades y no necesitan ser iguales al número de variables. Como se puede suponer, la transformación se hace de, alguna manera, por conveniencia.

Los *indicadores causales* son combinaciones lineales en las que el factor depende de las medidas (observables). En este caso, el factor hace las veces de la variable dependiente o criterio, en el sentido del análisis de regresión múltiple, mientras que las medidas actúan como variables independientes; por ejemplo, si las variables habilidad numérica y habilidad verbal integran el factor denominado aptitud académica, es de esperarse que los estudiantes que tengan una aptitud académica alta es porque su habilidad verbal y numérica es muy buena; no quiere decir esto que el hecho de tener una aptitud académica alta los convierta en habilidosos verbales y numéricos. El error, en consecuencia, está presente tanto en el factor como en las medidas (ver Bollen y Lennox, 1991 y Nunnally y Bernstein, 1995).

**Origen del análisis de factores.** Spearman (1904) ha sido considerado como el creador del AF. Él desarrolló este método con el propósito de probar su teoría de la inteligencia, en la cual sostenía que en todas las habilidades mentales subyacía un factor general que denominó  $g$ . Consideró que este factor era un tipo de norma mental de inteligencia y que esta sola norma era necesaria para explicar el área común existente entre todas las formas de diferencias individuales, en cuanto a las habilidades. En otras palabras, la teoría de Spearman suponía que cada fuente de diferencias individuales (pruebas) posee un factor único, pero como por definición, los factores únicos de diferentes pruebas no están correlacionados, todas las correlaciones entre las pruebas deben explicarse mediante un factor general. La teoría a veces se

denomina teoría de los dos factores de Spearman, ya que supone que cada prueba puede explicarse mediante un factor general y un factor único. Esta la solución bifactorial presupone que la varianza común entre las variables puede explicarse mediante dos tipos de factores comunes: un factor general y factores grupales. El método no establece previamente cuántos factores grupales, además del factor general, se requieren –el número de factores se determina por las correlaciones.

El análisis de factores como lo conocemos hoy día, fue introducido por Thurstone (1931) para una audiencia muy selecta de especialistas en el campo de la teoría de medición, la matemática y la estadística. Destacan, entre ellos, por su contribución al desarrollo del este método analítico autores tales como: Hotelling (1936); Holzinger y Harman (1941); Thurstone (1947); Cattell (1952); Fruchter (1954), Burt (1966); y Harman (1967).

Durante mucho tiempo el uso de este método estuvo restringido a los especialistas en álgebra lineal, particularmente en lo que se refiere a las operaciones con matrices (multiplicación e inversión de matrices), cuyo cálculo solía ser bastante laborioso mediante la utilización de una calculadora de mesa. Hoy día con el advenimiento del computador, los investigadores no necesitan ocuparse de estos cálculos complejos, sino más bien del uso de las posibilidades que les ofrece el método en sus diferentes aplicaciones. Sin embargo, en este caso, como en muchos otros en los que se utiliza el computador como herramienta de análisis de datos es necesario tener claro los tipos de análisis que se requieren y su justificación.

El usuario del computador, cuando trabaja con AF, debe saber que pedirle a la máquina que haga; por ejemplo, qué tipo de matrices de correlación calcular, qué método de factorización utilizar (método clásico, componentes principales, centroide, de raíz cuadrada), qué factores extraer

(rotados / no-rotados), qué método de rotación utilizar (varimax, quartimax, promax), cuántos factores rotar, qué pasa si se rotan muchos o pocos factores.

## Métodos de Factorización

La literatura recoge varios métodos de factorización, o de condensación, siendo algunos de los más conocidos los siguientes: el método clásico, el análisis de componentes principales, el método centroide, y otros. En este trabajo, sólo se considerarán los dos primeros.

### El Método Clásico

El método clásico de análisis de factores, se basa fundamentalmente en la creencia de que las correlaciones observadas en la matriz son principalmente el resultado de alguna regularidad subyacente en los datos. Más específicamente, se asume que la variable observada está influenciada por varias determinantes, algunas de las cuales son compartidas por otras variables en el conjunto; mientras que otras no son compartidas por ninguna otra variable. La parte de la variable que es influenciada por la determinante compartida es llamada usualmente *varianza común*; mientras que la parte que es influenciada por una determinante idiosincrática es llamada usualmente *varianza única*. Bajo este supuesto, la varianza única no contribuye a la relación entre las variables.

Del anterior supuesto se deriva también la noción de que las asociaciones observadas deben ser el resultado de las variables correlacionadas que comparten alguna de las determinantes comunes. Implícitamente se considera que las determinantes comunes no sólo explicarán toda las relaciones observadas en los datos, sino que además serán más pequeñas en número que las variables. El modelo del análisis de factores clásico puede ser expresado como sigue:

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + d_jU_j$$

$$j = 1, 2, \dots, n$$

Donde:

$z_j$  =  $j$ -ésima variable estandarizada

$F_i$  =  $i$ -ésimo factor hipotético

$U_j$  = factor único para la  $j$ -ésima variable

$a_{ji}$  = coeficiente de regresión múltiple estandarizado de la  $j$ -ésima variable en el  $i$ -ésimo factor

### **El Método de Componentes Principales**

El análisis de componentes principales es un método de transformación directa de un conjunto dado de variables dentro de un nuevo conjunto de variables compuestas o componentes principales, los cuales son ortogonales entre sí. No se requiere asumir ningún supuesto acerca de la estructura subyacente en los datos. El investigador debería interesarse sólo por determinar cuál es la mejor combinación lineal de las variables; mejor en el sentido de que una combinación lineal particular de variables sea capaz de extraer la mayor cantidad de varianza en los datos, como un todo, en comparación con cualquier otra combinación lineal de variables.

De acuerdo con lo anterior, el primer componente puede ser visto como el mejor resumen simple de combinación lineal exhibido en los datos. El segundo componente es definido, como la segunda mejor combinación lineal de las variables, bajo la condición de que el segundo componente sea ortogonal con respecto al primero. En consecuencia, el segundo componente debe explicar una proporción de varianza distinta a la explicada por el primer componente. Así, el segundo componente puede ser definido como la combinación lineal de variables que explica la mayor cantidad de varianza residual después que el efecto del primer componente es removido de los datos. Los componentes subsiguientes son definidos de manera análoga hasta

que toda la varianza es extraída de los datos. A menos que, como mínimo, una variable esté perfectamente determinada por el resto de las variables de la matriz de datos, la solución de componentes principales requiere tanto componentes como variables hay en los datos. El modelo estructural del análisis de componentes principales puede ser expresado como sigue:

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jn}F_n$$

Donde cada una de las **n** variables observadas es descrita linealmente en términos de **n** nuevos componentes no-correlacionadas:  $F_1, F_2, \dots, F_n$ . En vista de que cada componente es definido como el mejor resumen lineal de varianza dejada en los datos después que los componentes previos han actuado, el primer componente **m** puede explicar la mayor parte de la variación en los datos. Por un propósito práctico, el investigador retiene solamente los primeros componentes para posterior rotación.

## Tipos de Factores

Los factores reciben diferentes denominaciones de acuerdo con la manera como se organizan los coeficientes de correlación variable-factor (carga factorial) en la solución factorial final; así tenemos los siguientes tipos de factores:

- *Factor general*, cuando todas las medidas cargan en un solo factor, es decir, todas las medidas son salientes, en el sentido que las correlaciones variable-factor es igual o mayor que 0,50.
- *Factor de grupo*, es aquel en que sólo algunas variables son salientes, pero no todas.
- *Factor común*, es aquel que está integrado por variables que miden el mismo rasgo o dimensión, como ocurre en el caso del factor general y el factor de grupo.
- *Factor unipolar*, en donde todos los coeficientes tienen el mismo signo en el factor.
- *Factor bipolar*, cuando existen salientes con signos positivos y negativos en el mismo factor
- *Factor singular*, es aquel que está definido por una sola saliente.
- *Factor nulo*, cuando el mismo no tiene salientes.

A continuación se presenta un ejemplo en el que se ilustra cada uno de los tipos de factores antes descritos:

*Ejemplo 22*

<b>Variables</b>	<b>Factor I (general, unipolar)</b>	<b>Factor II (grupo)</b>	<b>Factor III (singular)</b>	<b>Factor IV (bipolar)</b>	<b>Factor V (nulo)</b>
X <sub>1</sub>	0,754	0,100	0,120	0,746	-0,006
X <sub>2</sub>	0,723	0,070	0,009	0,544	-0,120
X <sub>3</sub>	0,666	-0,050	0,207	0,657	0,009
X <sub>4</sub>	0,543	-0,001	0,110	0,543	0,207
X <sub>5</sub>	0,678	-0,120	0,100	0,678	0,110
X <sub>6</sub>	0,877	0,009	0,070	0,877	0,100
X <sub>7</sub>	0,746	0,207	0,050	-0,754	0,070
X <sub>8</sub>	0,544	0,110	0,001	-0,723	-0,050
X <sub>9</sub>	0,657	0,788	0,120	-0,666	-0,001
X <sub>10</sub>	0,557	0,878	0,009	-0,405	-0,120
X <sub>11</sub>	0,559	0,757	0,207	-0,508	0,009
X <sub>12</sub>	0,505	0,755	0,789	-0,500	0,207

## Aplicaciones del Análisis de Factores

La técnica del AF tiene múltiples aplicaciones en el quehacer científico, sin embargo, las más comunes son las siguientes: (a) como *técnica exploratoria*, con el propósito de examinar la estructura subyacente de un conjunto de datos, a fin de descubrir nuevos conceptos y una posible reducción de los datos; (b) como *técnica confirmatoria*, mediante la cual intentamos comprobar las hipótesis previamente asumidas acerca de la estructura subyacente en la matriz de datos, en términos del número de factores relevantes y cargas factoriales contenidas en la misma; este es el caso cuando utilizamos la técnica del AF como herramienta para determinar la validez de constructo de un instrumento de medición; y (c) como *técnica de medición*, cuando utilizamos los coeficientes de las cargas factoriales como ponderaciones de las variables de la matriz de datos a los fines de la construcción de índices que serán utilizados en análisis posteriores de una investigación.

### **El Análisis de Factores como una Técnica Exploratoria**

Suponga que un docente-investigador dispone de una serie histórica de datos en la que se incluyen las mediciones de un conjunto de variables de una muestra de estudiantes universitarios, las cuales son: comprensión de lectura, vocabulario, fluidez verbal, manejo de número, razonamiento matemático, velocidad de cómputo. El docente en referencia ha observado que los estudiantes que puntúan altos en estas medidas tienden a mantener un índice académico igual o superior al promedio de la muestra y desea explorar dicha base de datos con el propósito de determinar hasta donde estas variables son ortogonales o comparten alguna varianza común que hace posible

comprender su efecto en el rendimiento académico; es decir, se trata de identificar la estructura subyacente de la base de datos. Probablemente, de acuerdo con la literatura psicológica, el docente investigador va a identificar dos factores en los datos; uno verbal, formado por las tres primeras variables y otro numérico, integrado por las tres variables restantes.

### **El Análisis de Factores como una Técnica Confirmatoria**

Suponga que un investigador está desarrollando un instrumento para medir un constructo X, el cual hipotéticamente se expresa mediante los observables A, B y C. El instrumento ha sido diseñado de tal manera que los ítemes del mismo permiten operacionalizar dichos observables y se desea establecer la validez de constructo de dicho instrumento. En tal sentido, el investigador selecciona la muestra apropiada, administra el instrumento y analiza los datos con la técnica del AF. Algunas de las opciones de posibles resultados serían: (a) el AF indica que efectivamente existen tres factores bien definidos (A, B, C) tal como había sido hipotetizado por el investigador, en cuyo caso se comprobaría su hipótesis; (b) el AF indica que el instrumento está conformado por más factores de lo hipotetizado; (c) el AF señala que la estructura del instrumento contiene menos factores de los que originalmente se habían propuesto; o (d) el AF indica que los ítemes no comparten ninguna varianza común, es decir, no están midiendo el mismo constructo. En cualquiera de los tres últimos casos la hipótesis propuesta sería disconfirmada.

### **El Análisis de Factores como Medición**

Asuma que un investigador desea desarrollar un índice que le permita medir el constructo estrato socioeconómico (ESE) de una población determinada; al revisar la literatura especializada se encuentra que el constructo en referencia se expresa mediante los observables siguientes: educación del jefe de familia,

ocupación del jefe de familia, nivel de ingresos de la familia y tipo de vivienda. El investigador elabora un cuestionario y lo administra en una muestra representativa de la población; los datos son analizados factorialmente y se obtiene el resultado que se indica en el Cuadro 37.

*Cuadro 37. Resultados del Analisis Factorial*

<b>Variables</b>	<b>Factor I</b>
Educación ( $X_1$ )	0,756
Ocupación ( $X_2$ )	0,622
Ingresos ( $X_3$ )	0,825
Vivienda ( $X_4$ )	0,589

Los coeficientes obtenidos en las correlaciones de las variables con el factor pueden ser utilizados como los pesos relativos que cada variable tiene en la conformación del índice. Estos pesos hacen la misma función que los valores beta en el modelo de regresión múltiple. En consecuencia, cada coeficiente debe ser multiplicado por el valor obtenido por los sujetos en cada una de las variables, lo que permite obtener la expresión algebraica siguiente:

$$ESE = 0,756X_1 + 0,622X_2 + 0,825X_3 + 0,589X_4$$

## El Proceso del Análisis de Factores

En el proceso del AF se pueden distinguir cuatro fases fundamentales, a saber: crear la matriz de datos, calcular la matriz de correlación, extraer los factores iniciales (no-rotados) y rotar los factores iniciales hasta llegar a la solución final.

### La Matriz de Datos

La matriz de datos es una tabla de doble entrada formada por columnas y filas. En las primeras se ubican las *medidas*; mientras que en la segunda se identifican los *sujetos*, cuyos atributos han sido medidos. En las celdas que se forman al cruzarse las columnas con los sujetos están ubicadas las puntuaciones; así, por ejemplo, la celda ***a*1** representa la puntuación del sujeto **1** en la medida o variable ***a***, en el ejemplo de la matriz de datos que aparece a continuación:

*Ejemplo 23*

		Variables				
		a	b	c	...	k
Sujetos	1	$a_1$	$b_1$	$c_1$	...	$k_1$
	2	$a_2$	$b_2$	$c_2$	...	$k_2$
	3	$a_3$	$b_3$	$c_3$	...	$k_3$
	4	$a_4$	$b_4$	$c_4$	...	$k_4$
	5	$a_5$	$b_5$	$c_5$	...	$k_5$
	6	$a_6$	$b_6$	$c_6$	...	$k_6$
	7	$a_7$	$b_7$	$c_7$	...	$k_7$
	.	.	.	.	...	.
	.	.	.	.	...	.
	.	.	.	.	...	.
	N	$a_n$	$b_n$	$c_n$	...	$k_n$

### Las Medidas

El término medidas se refiere a cualquier conjunto de atributos que puedan cuantificarse y ser registrados mediante la utilización de tests psicológicos, escalas, pruebas de rendimiento académico, variables fisiológicas, entre otras. Por su parte, el término sujetos se refiere a cualquier clase de objetos de los cuales se obtienen medidas. Es necesario que haya, al menos una puntuación para cada sujeto en cada medida y se requiere que la matriz sea más larga que ancha; es decir debe contener más sujetos que medidas; en caso contrario, el AF operaría al azar.

### La Matriz de Correlaciones

Como ya se ha mencionado, después de seleccionar la base de datos, la fase

siguiente en el AF es calcular la matriz de asociación entre las variables relevantes del conjunto de datos. Dicha matriz puede estar conformada de dos maneras: a partir de la relación entre las variables (o atributos), que es la más común y se conoce con el nombre de matriz tipo-R; o de la asociación entre los sujetos (u objetos), en cuyo caso recibe el nombre de matriz tipo-Q.

*Ejemplo 24. Matriz de Correlaciones*

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
$X_1$	1.00	0,75	0,83	0,32	0,28	0,36
$X_2$	0,75	1.00	0,70	0,25	0,31	0,32
$X_3$	0,83	0,70	1.00	0,39	0,25	0,33
$X_4$	0,32	0,25	0,39	1.00	0,79	0,82
$X_5$	0,28	0,31	0,25	0,79	1.00	0,76
$X_6$	0,36	0,32	0,33	0,82	0,76	1.00

Como se puede observar, la matriz tiene dos parte iguales separadas por una diagonal cuyos valores son igual a uno. Asimismo, se puede apreciar que las variables  $X_1$ ,  $X_2$  y  $X_3$  forman un primer grupo de correlaciones salientes que probablemente darán origen a un factor; mientras que las variables  $X_4$ ,  $X_5$  y  $X_6$ , forman un segundo grupo de intercorrelaciones que pareciera integrarían un segundo factor. Obsérvese, también, que todos los signos de los coeficientes son positivos y que sus magnitudes varían.

**La Extracción de los Factores Iniciales**

En esta fase se trata de explorar la posibilidad de reducir los datos mediante la construcción de un conjunto de nuevas variables sobre la base de la intercorrelaciones exhibidas en los datos. En tal sentido, las nuevas variables

pueden ser definidas como una transformación matemática exacta de los datos originales o pueden ser vistas como inferencias sobre la estructura hipotética asumida en la estructura de los datos y su fuente de variación. El primer caso, en que los factores son previamente definidos, recibe el nombre de Análisis de Componentes Principales (ACP); mientras que el segundo, en donde los factores son inferidos, se denomina Análisis de Factores Clásicos (AFC). No obstante, que los factores iniciales sean definidos o inferidos, generalmente son extraídos de tal manera que un factor es independiente de los otros; es decir los factores son ortogonales.

*Ejemplo 25. Matriz Factorial No Rotada*

	Factor I	Factor II	$h^2$
$X_1$	0,77	0,54	0,88
$X_2$	0,72	0,52	0,79
$X_3$	0,76	0,52	0,85
$X_4$	0,79	-0,51	0,88
$X_5$	0,74	-0,54	0,84
$X_6$	0,79	-0,48	0,85
% de Varianza	0,58	0,27	0,85

La matriz factorial (no rotada) anterior ha sido calculada a partir de la matriz de correlación presentada en la sección anterior. Los coeficientes de los factores I y II pueden ser considerados ya sea como elementos del patrón (pesos de regresión) o como elementos estructurales (correlaciones), denominadas frecuentemente como cargas factoriales. Obsérvese que el factor I es de tipo general, mientras que el II es bipolar. La variable  $X_1$  puede

ser estimada como 0,77 veces el factor I más 0,54 veces el factor II. La variable  $X_1$  también se correlaciona 0,77 y 0,54 con los factores I y II. La columna de la extrema derecha representa la comunalidad de las variables, que se obtiene (en una solución ortogonal) sumando los cuadrados de los elementos estructurales (o patrón), por ejemplo, en el caso de la variable  $X_1$ ,  $h^2 = 0,88 (0,77^2 + 0,54^2)$ . Este valor de la comunalidad puede ser interpretado como la proporción de varianza común que una determinada variable explica en los factores con los cuales está correlacionada. Asimismo, la  $h^2$ ) puede ser vista como la correlación múltiple cuadrada ( $R^2$ ) al predecir la variable a partir de los factores.

Las primeras dos columnas en la última fila contiene información sobre la varianza factorial o la cantidad de varianza explicada para cada factor; éstas se obtienen al promediar los elementos estructurales (o patrón) sobre el número de variables.

### **Rotación de Factores hasta la Solución Final**

No existe una manera única de llegar a una solución factorial final, entre otras razones, porque la configuración de la estructura de los factores depende del tipo de rotación realizada la cual, a su vez, está asociada a diferentes procedimientos estadísticos, sin que se violen los supuestos matemáticos básicos de la técnica del AF. Además, no todas las soluciones de factores estadísticos son igualmente significativos desde el punto de vista teórico. Algunos procedimientos son más laboriosos e informativos que otros y cada uno dice algo un poco diferente acerca de la estructura de los datos. De allí que el investigador tenga que escoger el mejor método de rotación para alcanzar la solución final que puede satisfacer las necesidades teóricas y prácticas del problema de investigación.

Por lo general, el investigador debe optar entre un método de

rotación ortogonal (factores no-correlacionados) o un método de rotación oblicuo (factores correlacionados). El criterio básico para decidir cual de los dos utilizar, en un momento dado, es el mismo: cómo reducir los datos a dimensiones estructurales más pequeñas que tengan el mayor significado teórico. Por lo general, los factores ortogonales son matemáticamente más simples de manejar; mientras que los factores oblicuos son más realistas. Sin embargo, no se podría decir, que un método ofrezca algún tipo de ventaja sobre el otro; por lo tanto, la decisión de utilizar uno u otro depende de las necesidades particulares de un problema dado de investigación.

*Ejemplo 26*

	Factor I	Factor II	$h^2$
$X_1$	0,93	0,16	0,89
$X_2$	0,78	0,16	0,63
$X_3$	0,86	0,18	0,77
$X_4$	0,20	0,90	0,85
$X_5$	0,16	0,84	0,73
$X_6$	0,23	0,86	0,79
% de Varianza	0,39	0,39	0,78

Como se puede apreciar en la matriz factorial anterior (rotación varimax), las cargas factoriales en cada factor han quedado perfectamente definidas por sus salientes respectivos. Esta matriz presenta una estructura simple, cuya interpretación se hace más sencilla en comparación con la matriz no-rotada.

**Métodos de rotación.** Algunos de los métodos analíticos de rotación más

conocidos son: Varimax, Cuartimax y Promax. El *método Varimax* fue desarrollado por (Kaiser, 1958; Gorsuch, 1966) para eliminar los factores generales, ya que su interés está en capturar muy bien el significado de la estructura simple dentro de un marco de referencia ortogonal; opera sobre las columnas de la matriz, para lo cual normaliza previamente las variables.

El *método Cuartimax* fue desarrollado por Neuhaus y Wrigley (1954); es ortogonal, opera sobre las filas o hileras de la matriz de datos y su interés primordial está en identificar un factor común como parte de la estructura subyacente en la matriz de correlaciones. Este criterio se satisface cuando todas las variables se correlacionan altamente con un mismo factor. Finalmente el *método Promax* permite lograr una solución analítica oblicua (no ortogonal), fue desarrollado por Hendrickson y White (1964) y su objetivo es similar al de varimax en que busca maximizar la propagación (varianza) de los elementos del patrón en un factor.

De acuerdo con Nunnally (1987), las supuestas ventajas de las rotaciones oblicuas son más conceptuales que matemáticas. En la práctica existen argumentos que favorecen tanto las rotaciones ortogonales como las oblicuas, ya que ambas son matemáticamente legítimas, por lo tanto el uso de una u otra es cuestión de gusto o de interés personal del investigador.

***Criterios de rotación.*** ¿Cuál es el propósito de rotar los factores? En principio, desde el punto de vista estadístico, los factores rotados son tan adecuados como los no rotados. No obstante, la razón fundamental que existe para rotar los factores es buscar una solución fácilmente interpretable (Nunnally, 1987). Una regla general de la rotación es que ésta debe hacerse de tal forma que cada variable tenga carga sobre uno y sólo uno de los factores. Ello no implica que haya la misma cantidad de variables puras para todos los factores, sólo se requiere que haya algunas para cada factor. La

situación ideal en estos casos condujo a Thurstone (1947) al concepto de *estructura simple*. Normalmente una matriz rotada tiene una estructura más simple que una no rotada. Lo que normalmente se busca en la rotación es que hayan algunas variables relativamente puras en cada factor.

***Determinación del número de factores.*** Existen varios criterios para definir el número de factores a ser extraídos en una solución factorial. Uno de los más comunes es el conocido como la regla Kaiser-Guttman (Guttman, 1954; Kaiser, 1960, 1970), según la cual se deben seleccionar factores con un auto valor igual o mayor que 1 . La regla requiere que un componente explique, al menos, tanta varianza como una variable individual. Sin embargo, más allá de esta consideración estadística, es importante que la solución factorial adoptada siempre tenga algún sentido teórico. Los factores de interés describen aspectos que las variables tienen en común, para lo cual se requiere, como mínimo, que dos variables definidas aparezcan como salientes en los factores rotados.

Algunos autores consideran que al extraer demasiados factores se diluye la estructura de los factores rotados, se reduce el número de salientes y, por lo tanto, hace los factores resultantes más difíciles de interpretar. Agregar un componente siempre aumenta la varianza explicada, pero disminuye la carga estructural promedio. Por otra parte, se pueden perder facetas sutiles pero potencialmente importantes de los datos si se retienen muy pocos factores (ver Nunnally y Bernstein, 1995). En resumen, se puede señalar que no existe un regla fija para decidir sobre el número de factores que deben ser extraídos como resultados de una solución factorial final. Es necesario tomar en cuenta las consideraciones estadísticas, pero sin dejar de lado la relevancia teórica, lo cual debe llevar a una interpretación, con sentido, de los factores resultantes.

A manera de resumen se presentan a continuación la secuencia de etapas y opciones que son necesarias considerar para llevar a efecto un AF (ver Cuadro 38).

*Cuadro 38. Cuadro resumen del análisis de factores*

<b>Etapas el Análisis de Factores</b>	<b>Opciones</b>	<b>Referencias</b>
Matriz de datos	Única (sujetos por medidas)	Matriz X
Matriz de correlación	Correlaciones entre medidas	Matriz tipo-R
	Correlaciones entre unidades	Matriz tipo -Q
Extracción de factores Iniciales	Factores definidos	ACP
	Factores inferidos	AFC
Rotación hasta la solución final	Factores no-correlacionados	Rotación ortogonal
	Factores correlacionados	Rotación oblicua

*ACP: Análisis de Componentes Principales*

*AFC: Análisis de Factores Clásico*

## Análisis de Factores y Desarrollo de Instrumentos

Como ya se mencionó anteriormente, el AF tiene diferentes aplicaciones en la investigación social, psicológica y educativa. Una de ellas se presenta en el campo de la teoría de medición, particularmente en lo que respecta al desarrollo de instrumentos, tipo escala de estimación, pruebas de habilidades, aptitudes, conocimientos y otras. Una vez que ha sido concluida la primera versión, en el proceso de desarrollo del instrumento, una de las tareas subsiguientes consiste en realizar un análisis de ítem, para conocer el nivel de funcionamiento de los reactivos y determinar cuáles pueden cumplir mejor su propósito.

### **Análisis de Ítem**

Existen diferentes procedimientos convencionales que permiten cumplir con las diferentes tareas que supone realizar un análisis de ítem, por ejemplo la contrastación del desempeño de grupos extremos mediante una prueba *t* o mediante una correlación biserial ítem-total.

Un propósito similar puede ser cumplido por el AF, tomando en cuenta las correlaciones ítem-factor (carga factorial), ya que ellos son análogos al coeficiente de validez biserial en el análisis de ítem. De esta manera se podrán identificar ítems con altas o bajas cargas factoriales; por ejemplo, coeficientes igual o mayor que 0,50 se consideran adecuados para integrar la versión final del instrumento, así como para evaluar la estructura de un factor (Guertin y Bailey, 1970). Es altamente recomendable eliminar aquellos ítems que tengan cargas factoriales inferiores al criterio antes mencionado a fin de proceder a correr nuevamente el AF con los ítems seleccionados.

Convencionalmente, el proceso de análisis de ítem supone la relación de

cada ítem con un criterio, el cual es, por lo general, la puntuación total del instrumento, resultante de la sumatoria de las calificaciones obtenidas por los sujetos al responder los reactivos. Este procedimiento asume implícitamente que se trata de un instrumento unidimensional, lo cual no es necesariamente cierto, ya que el puntaje total de una prueba puede integrar de manera desproporcionada diferentes factores que no son aparentes.

Suponga, por ejemplo, que usted administra una prueba de 60 ítemes a un grupo de estudiantes de séptimo grado. La prueba está formada por 25 reactivos de cálculo aritmético y 35 preguntas de habilidad verbal, y después de administrado el instrumento se suman las puntuaciones para expresar el desempeño de los estudiantes. Puesto que los 35 ítemes verbales son más numerosos que los 25 de aritmética, ellos contribuirán más al puntaje total. En otras palabras, la calificación total dependerá más del factor verbal, que del aritmético. En un análisis de ítem los reactivos de cálculo aritmético tendrán coeficientes de validez más bajos y, en consecuencia, serán eliminados más frecuentemente que los ítemes verbales.

El ejemplo anterior ilustra bastante bien cuán cuestionable puede ser el puntaje total de un instrumento cuando se utiliza como su criterio de validación. De allí que el análisis de ítem debería ser usado cuando ya se tenga una estructura factorial establecida y se esté buscando un conjunto de ítemes relativamente puro para medir un rasgo determinado. Si un conjunto de ítemes parecieran cargar en factores similares en varios estudios y pareciera medir el rasgo que se busca, los mismos podrían ser integrados y obtener un puntaje factorial total a fin de utilizarlo como un criterio para correlacionar los ítemes.

Posteriormente, vendría el procedimiento de refinamiento, eliminando los ítemes que tengan coeficientes bajos. Algunos autores consideran que el

análisis de ítem debería venir sólo en esta etapa de refinamiento del instrumento (ver Guertin y Bailey, 1970). Ciertamente, ello no es adecuado en las etapas tempranas del desarrollo de un instrumento cuando hay interés en identificar rasgos. El AF de una prueba arrojará varios factores, tal vez rasgos o componentes de logro, dependiendo de lo que se esté midiendo con el instrumento.

### **Estudio de la Confiabilidad**

Un índice de confiabilidad mínima para los ítems de un test es basa en la comunalidad encontrada en el proceso de factorización. El coeficiente de confiabilidad test-retest de un ítem nunca debería ser inferior a la raíz cuadrada de su comunalidad. Una baja comunalidad no es sinónimo de baja confiabilidad; sin embargo, una alta comunalidad si asegura una alta confiabilidad.

Normalmente, la comunalidad de los ítems de los cuestionarios promedia 0,40 o superior, dependiendo de la naturaleza de las preguntas, las características de los sujetos y la naturaleza del proceso de recolección de datos. La raíz cuadrada de 0,40 es aproximadamente 0,63, que es una confiabilidad adecuada para un simple ítem. Los ítems se juntan para obtener subescalas que son más confiables que los ítems individuales, debido a la mayor extensión de la subescala. Cuando se combina cuatro o cinco ítems con una confiabilidad de 0,60, tendremos una medida confiable del atributo que se está intentando medir.

### **Estudio de la Validez**

El AF constituye un recurso de gran utilidad heurística en el estudio de la validez, cuando se desarrollan instrumentos de medición. El estudio de la validez implica evaluar el grado en que un instrumento se aproxima, desde el punto de vista métrico, al objeto que pretende medir, es decir, se intenta

averiguar cuál es su nivel de exactitud con que cumple su propósito. En otras palabras, cuando se estima la validez de un test, se quiere saber que rasgo se desea que mida, el cual se conoce como variable de criterio. Interesa saber que tan bien corresponden las posiciones de los individuos en la distribución de los puntajes obtenidos a sus posiciones en el continuo que representa la variable criterio (Magnusson, 1983).

La validez es usualmente estimada mediante la correlación de una variable en estudio con un criterio, el valor obtenido recibe el nombre de coeficiente de validez, el cual indica la relación que hay entre los datos obtenidos en el test y las puntuaciones que se utilizan, con un grado conocido de certeza, como índices para las calificaciones del individuo en la variable de criterio. Un procedimiento similar puede ser utilizado mediante la utilización del AF para verificar los diferentes tipos de validez (constructo, predictiva y contenido).

El estudio de la validez está en el centro de la construcción misma del conocimiento científico, ya que de alguna manera la comprobación de la validez del conocimiento es el puente que conecta la teoría con la realidad, particularmente en lo que se refiere a la validez de constructo.

Explicar los constructos consiste principalmente en determinar (a) la estructura estadística interna de una serie de variables que se dice miden un constructo; y (b) las estructuras cruzadas entre las diferentes medidas de un constructo y aquellas de otros constructos. El AF se utiliza de manera directa al abordar ambas cuestiones (Nunnally y Bernstein, 1995).

Suponga que sobre la base de una teoría determinada se ha desarrollado un instrumento para medir el constructo X, el cual ha sido hipotetizado como integrado por dos dimensiones: A y B. El instrumento se operacionaliza mediante la incorporación de 20 ítemes cada una. Si todos los reactivos de la

dimensión A tienen una alta correlación entre sí, de la misma manera que en la dimensión B; esto quiere decir que es esperable una alta correlación entre los ítemes y su dimensión correspondiente. Esta evidencia indica que las dos dimensiones del constructo en referencia tienen una estructura interna fuerte. La estructura cruzada de las dos dimensiones sería apoyada, además, si los dos factores se correlacionan de manera sustantiva. De esta manera, la validez de constructo queda verificada mediante la técnica del AF confirmatorio.

El AF también proporciona evidencia útil acerca de la validez de contenido. Nunnally y Bernstein (1995) han desarrollado un ejemplo que ilustra esta posibilidad. Suponga que se realiza un análisis factorial en una batería de pruebas de rendimiento académico y que la presunta prueba de matemática se correlaciona altamente con un factor de comprensión verbal. Esto implica que la redacción de los reactivos fue excesivamente difícil. Tal información aportada por el AF llevaría a una nueva redacción de los reactivos de la prueba, a fin de eliminar el problema.

Por otra parte el AF es importante en la selección de instrumentos que serán tratados como predictores. En lugar de construir una prueba nueva para cada problema que surja, se debería seleccionar un instrumento predictivo de un banco de instrumentos disponibles. El AF permitiría evaluar previamente la composición factorial de dichos instrumentos a fin de utilizar los factores indentificados como predictores. Al respecto, Nunnally y Bernstein (1995) plantean que “es mucho más fácil formular hipótesis acerca del posible poder predictivo a partir de factores particulares, que formular hipótesis acerca del poder predictivo de instrumentos desarrollados sobre una base ad hoc” (p. 124).

La validez de un instrumento, en el contexto del AF, también puede ser

chequeada mediante el procedimiento de validación cruzada, a fin de determinar la estabilidad de la estructura factorial encontrada en el análisis de los ítems de un instrumento. En este caso, en lugar de aplicar el instrumentos dos veces en muestras distintas para comparar y verificar si las estructuras factoriales obtenidas se corresponden, la muestra general (por ejemplo  $n = 400$ ), se divide en dos submuestras ( $n_1 = n_2 = 200$ ) se factorializan y se comparan sus estructuras factoriales.

### **La Importancia de la Muestra**

Uno de los supuestos que se asumen cuando se utilizan instrumentos desarrollados para una población diferente es que ellos, al igual que los integrantes de nuestra muestra, son seres humanos y, por lo tanto, son todos similares. En consecuencia, se emplea el mismo instrumento para calcular subescalas, puntuaciones totales y comparar diferencias en los promedios, pero en muchos de estos casos tales procedimientos son incorrectos. Con frecuencia los ítems que se integran en una muestra de normalización pudiera no ser válida en otra muestra en particular. Es necesario que las muestras de normalización incluyan sujetos heterogéneos a fin de favorecer la variación en la variable de estudio. A mayor heterogeneidad en la muestra, más clara será la estructura factorial obtenida. Las correlaciones altas incrementan la carga factorial, una mejor determinación en la localización del factor y clarificación de la estructura factorial.

Un instrumento no logra una estructura factorial sin referencia a una muestra en particular. Pudiéramos obtener algo que tenga muy poco valor práctico o significado si no se selecciona una muestra apropiada (representativa). Pero ¿cuánto sujetos deberían integrar una buena muestra, a fin de lograr una medida confiable? Al respecto no hay una regla fija, ya que ello depende, por ejemplo, del propósito de la investigación, la naturaleza de

las variables, la confiabilidad de la medida, etc. Por lo general, se recomienda seleccionar muestras no muy grandes (200 a 300 sujetos), en una primera etapa, ya que básicamente interesa obtener una estructura factorial preliminar. Al respecto, es importante recordar que entre más sujetos se utilicen, más confiables serán los coeficientes de correlación y más estable será la estructura factorial. El coeficiente de confiabilidad aumenta a medida que se incrementa la raíz cuadrada del número de casos. No obstante, la ganancia en precisión mediante el incremento del tamaño de la muestra no es lineal.

En resumen, en este trabajo se presentó una síntesis del método AF haciendo énfasis en sus implicaciones teóricas, metodológicas y prácticas para los investigadores de las ciencias sociales, psicología y educación. Se destacó la aplicación del método en su aspecto confirmatorio, el cual es de gran utilidad en el campo de la teoría de medición, particularmente en lo que respecta al desarrollo de instrumentos. Se analizaron las cualidades analíticas y heurísticas del método en la construcción y validación de teorías científicas. Se concluye que el AF es un método que ofrece ventajas comparativas, cuando se trata de establecer las características técnicas de un instrumento a fin de apoyar la toma de decisiones apropiadas para su mejoramiento. En tal sentido, se recomienda ampliamente la utilización del método de AF en el proceso de desarrollo de instrumentos de medición.

PART 2

PARTE II TÉCNICAS DE  
ANÁLISIS CUALITATIVO

# Observación Participante y Entrevista en Profundidad

## Observación Participante

La observación, en el contexto de la psicología cognitiva de procesamiento de información, se considera como el proceso mediante el cual las personas obtienen información directa o indirectamente de su entorno, mediante los receptores sensoriales. Consiste en la identificación permanente de las características de las cosas en la interacción del sujeto con su ambiente. En este tipo de identificación se pueden apreciar dos momentos: (a) el contacto con el objeto o situación a nivel concreto; y (b) la abstracción del objeto o situación, para transformarlo en una imagen o representación mental (Sánchez, 1991). Por ejemplo, en el plano físico, un sujeto puede identificar las características externas del objeto mesa y compararlo con otros objetos similares, aislar los atributos que le son comunes y, de esa manera, llegar al concepto de mesa el cual, como todo concepto, constituye una abstracción.

La observación en cuanto técnica de investigación social puede ser entendida como un procedimiento, intencional y sistemático, que permite obtener datos y/o información sobre una situación tal cual como ésta ocurre, sin modificarla o manipularla, con el propósito de explicarla, comprenderla o describirla. Bunge (2000) distingue cinco elementos en el proceso de observación, ellos son: (a) el observador y sus características personales, profesionales y culturales; (b) el objeto de la observación, que es la realidad objeto de estudio; (c) las circunstancias de la observación o las condiciones concretas que rodean al hecho de observar y que terminan por formar parte de la propia observación; (d) los medios de la observación: son los sentidos y los instrumentos de registro de datos e información; y (e) los resultados obtenidos, entendidos como el conjunto de saberes debidamente

estructurados a fin de ser integrados a un cuerpo más amplio de conocimientos.

Las fases que se siguen en el uso de la observación como técnica de investigación son las siguientes: (a) definir y delimitar la situación objeto de estudio, (qué se va a observar); (b) establecer los objetivos de la observación; (c) determinar la forma con que se van a registrar los datos (cuaderno de notas, registros anecdóticos, fichas, grabaciones de audio e imagen, fotografías); (d) apreciar cuidadosa y críticamente los elementos de la situación objeto de estudio y sus interacciones; (e) registrar los datos observados; (f) analizar e interpretar los datos; y (g) elaborar conclusiones sobre lo observado.

Ander-Egg (1995) ha clasificado la técnica de la observación de la siguiente manera:

- **Observación estructurada.** Es aquella en la que se establece previamente los aspectos que se quiere observar y se recurre a instrumentos para la recopilación de datos o hechos observados.
- **Observación no estructurada.** Consiste en reconocer y anotar los hechos sin recurrir a la ayuda de técnicas especiales, el peligro de este tipo de observación es que se puede presentar como una sensación de que sabemos más de lo que hemos visto.
- **Observación participante.** Consiste en examinar al grupo por dentro y por fuera, esto es, por fuera captando lo que el individuo o el grupo quieren decir y quieren dejar ver, y por dentro cuando el grupo acepta y se acostumbra a ver al observador y actúa como si él no estuviera.

- **Observación no participante.** Se trata de una observación con propósitos definidos. El investigador se vale de ella para obtener información y datos sin participar en los acontecimientos de la vida del grupo que estudia, permaneciendo ajeno al mismo.
- **Observación individual.** Es aquella que realiza una sola persona. Al respecto, es importante evitar que el investigador se proyecte sobre lo observado con el riesgo de provocar distorsiones.
- **Observación en equipo.** Permite amplias las posibilidades de control, gracias a que todos observan la misma situación, cada uno atiende un aspecto diferente y se construye una red de observadores.
- **Observación de la vida real.** Nos permite captar los hechos tal y como se presentan, sin preparación, las ciencias sociales utilizan esta forma de observación llamada también de campo.
- **Observación de laboratorio.** Las ciencias experimentales recurren a ésta, lo que permite al investigador pasar al “caso concreto” al “caso típico” para descubrir las condiciones de producción del fenómeno a través de hechos de significación general.

En este capítulo nos interesa distinguir, de manera particular, las modalidades de observación que se derivan de los enfoques de investigación cuantitativo y cualitativo. Así, la observación en el primer caso se caracteriza por ser altamente estructurada, sistemática y controlada; los investigadores buscan observaciones objetivas y confiables a partir de las cuales puedan realizar inferencias válidas. Kerlinger y Lee (2002) consideran que bajo este

enfoque “los investigadores tratan la observación del comportamiento como parte de un procedimiento de medición: asignan valores numéricos a objetos de acuerdo con reglas, en este caso, de acuerdo con actos o secuencias de actos de comportamiento humano” (p. 662).

Por su parte, la observación bajo el enfoque cualitativo a diferencia de la anterior, es flexible, abierta y no controlada; el investigador está interesado en apreciar la situación bajo estudio en términos de las acciones, interacciones y transacciones entre los actores del hecho observado, con el propósito de comprender, describir e interpretar lo ocurrido. Estas características de la observación cualitativa no significan que la misma se realice sin orden, rigor metodológico o falta de objetivos. Esta flexibilidad más bien obedece a la visión de la temática que involucra al investigador como sujeto social y a los objetivos propuestos así como a los procesos interpretativos que hará sobre los resultados de la información obtenida.

La observación cualitativa puede ser de dos tipos: participante y no-participante. En la primera, el investigador se integra activamente al grupo social u organización que desea estudiar y desde adentro se comporta como un miembro más; interacciona con los demás actores al mismo tiempo que registra sistemáticamente los datos e información que estima son relevante en función de los objetivos de la investigación. Woods (1987) considera que “la observación participante es un medio para llegar profundamente a la comprensión y explicación de la realidad por la cual el investigador «participa» de la situación que quiere observar, es decir, penetra en la experiencia de los otros, dentro de un grupo o institución. Y pretende convertirse en uno más, analizando sus propias reacciones, intenciones y motivos con las de los demás... Debe intentar combinar la profunda implicación personal con un cierto distanciamiento” (p. 50). Por su parte, en

la observación no-participante, el investigador es un espectador interesado y motivado por estudiar en profundidad lo que ocurre en el grupo u organización bajo estudio, sin intervenir directamente en la situación o el hecho observado.

En el uso de esta técnica de la observación participante, el investigador es un observador declarado y aceptado por la comunidad. La premisa básica de este tipo de investigación es la convivencia en el medio para el conocimiento del mismo. Es una manera de investigación en profundidad y ocupa por lo general un tiempo prolongado. Esta técnica, aunque muy vinculada inicialmente a los estudios etnográficos de grupos étnicos y otros grupos sociales minoritarios, con el tiempo se ha generalizado su uso a otros contextos, incluyendo el educativo. De acuerdo con Guasch (1997), “en la observación participante lo que se pretende es que el investigador se convierta él mismo en un ‘*nativo*’ a través de la inmersión en la realidad social que analiza o, dicho de otro modo, lo que el investigador pretende es aprehender y vivir una vida cotidiana que le resulta ajena. Para ello se ocupa de observar, acompañar y compartir con los actores las rutinas típicas y diarias que conforman la experiencia humana” (p. 1). Al aplicar esta técnica, el investigador debe determinar qué es lo que quiere observar y cómo va a registrar sus datos. Debe plantear su estrategia anticipadamente, así como establecer listas y registros de observación de manera que la observación sea selectiva, concentrándose ésta en los detalles sustantivos.

Por otra parte, el investigador debe estar consciente que al insertarse en la comunidad o grupo objeto de estudio debe asumir las normas y reglas de funcionamiento del grupo y actuar conforme a ellas. No obstante, el observador no actuará de manera neutra sino que en su acción estará reflejado todo su repertorio sociocultural previo, experiencia profesional y supuestos

teóricos. La principal ventaja de esta situación es que el observador adquiere datos e información derivados directamente de la situación bajo estudio; mientras que la principal desventaja es el riesgo de que el investigador pudiera exagerar un poco uno de sus roles, por ejemplo, tener más injerencia en las actividades del grupo (participante) que en su rol de apreciar la realidad y registrar datos e información relevante (observador).

## Momentos de la Observación

En esta sección consideraremos los cuatro momentos básicos del proceso de la observación participante, como son: la preparación para la entrada al campo, desarrollo de la observación en el escenario, la salida del escenario y el análisis de datos e información.

**Preparación.** Una de las consideraciones previas que debe hacer el observador es precisar si su propósito es de tipo sustantivo o teórico. El primero se refiere a interrogantes relacionadas con problemas específicos en un particular tipo de escenario (por ejemplo, estudiar los problemas de una escuela). El segundo está más estrechamente relacionado con problemas sociológicos básicos tales como la socialización, la desviación y el control social. Estos conceptos son análogos a los utilizados por Glaser y Strauss sobre teoría sustantiva y teoría formal (Taylor y Bogdan, 1996).

Una segunda consideración se refiere al acceso al escenario u organización. Por lo general, ello se logra mediante un trámite formal ante las autoridades competentes. En el caso de un docente-investigador que intente aplicar la investigación-acción en su aula de clase, tendría que solicitar la autorización ante el director del plantel o ante el supervisor del distrito escolar. En cualquier caso, el docente-investigador deberá justificar las ventajas que tendrá el estudio para el mejoramiento de la calidad del proceso de aprendizaje que se desarrolla en el aula de clase, enfatizando que en ningún caso la investigación interferirá con el logro de los objetivos instruccionales del aula.

Finalmente, el investigador debe tener claro los objetivos de la observación en el marco de una planificación flexible que dé respuesta a

preguntas tales como las siguientes: ¿qué investigar?, por ejemplo, la cultura escolar, la interacción en el aula de clase, las actitudes de los docentes hacia las tecnologías de la información y la comunicación; ¿cómo observar?, directamente en el escenario donde se dan los hechos, prestando atención a todos los aspectos que parezcan relevantes, interaccionando con los actores, acudiendo a documentos, participando en las actividades de rutina; ¿dónde observar?, por ejemplo, en el aula de clase, en la institución escolar o académica como un todo; ¿qué observar?, el comportamiento de los actores, los hechos relevantes que ocurren en el grupo, la cultura de la organización, ¿cuándo observar?, todo depende del horario de funcionamiento del grupo y del cronograma del investigador; pudiera ser a una determinada hora, en determinado turno o todos los días, hasta que sea necesario; ¿cómo registrar?, existen varios tipos de registros, tales como el narrativo (diario o notas de campo) y los audiovisuales (fotografías, grabadores de audio, vídeos...), ¿cómo analizar? Existen diferentes tipos de análisis de la información textual, que pueden ser utilizados tanto en forma manual como por procedimientos electrónicos, como son el análisis del contenido, análisis del discurso, análisis estadístico de datos textuales y el análisis de datos cualitativos (Armony, 1997).

***Desarrollo de la observación en el escenario.*** Algunos autores como Taylor y Bogdan (1996) creen que es muy importante recordar todo aquello que el investigador ve, oye, o siente mientras está en el campo. Por ello recomiendan tener en cuenta los aspectos siguientes: prestar atención, pasar de una visión amplia del contexto a otra más reducida, buscar palabras claves en las observaciones de la gente, concentrarse en las observaciones primeras y últimas de cada conversación, reproducir mentalmente las observaciones y escenas, abandonar el escenario en cuanto haya observado todo lo que esté en

condiciones de recordar, tomar notas tan pronto resulte posible, después de la observación, dibujar un diagrama del escenario y trazar sus movimientos en él, grabar conversaciones y acontecimientos.

A diferencia de lo que ocurriría en cualquier estudio de naturaleza antropológica o sociológica, donde el investigador tendrá que hacer una serie de consideraciones previas para la entrada al escenario, en el caso del contexto educativo este paso se cubre con relativa normalidad una vez que se tiene la autorización para realizar el estudio y más aún si el investigador es el propio docente.

Es importante tener presente que una vez que se entra al escenario, además de la observación como tal, el investigador puede disponer de otras fuentes de información complementarias tales como entrevistas informales y análisis de documentos, entre otros. En el caso de las entrevistas, es necesario pensar en quienes pueden ser los informantes clave. Por ejemplo, si se tratase de un estudio etnográfico en un barrio de una gran ciudad, el investigador podría entrevistar a líderes comunales así como personas comunes y corrientes, de ambos géneros, de diferentes edades. Sin embargo, si se tratara del uso de la observación participante en una institución educativa, el tipo de informante a entrevistar dependería de la naturaleza y propósito de la investigación. Los informantes potenciales en este tipo de estudio suelen ser el personal directivo, miembros del personal docente, los propios alumnos y los padres y representantes.

En cuanto a la utilización de documentos como fuente de información adicional es altamente recomendable el tener acceso a informes oficiales, archivos, correspondencias, contratos, nóminas de salarios, formularios de evaluación y diarios. Asimismo es valioso utilizar documentos históricos y públicos a fin de obtener una perspectiva más amplia respecto de la situación

bajo estudio. Estas fuentes adicionales de información permitirán ampliar y profundizar en el conocimiento de determinados temas que el investigador considere necesario para comprender mejor la situación bajo estudio. Asimismo, el investigador podrá usar posteriormente estas fuentes como un recurso metodológico para la triangulación de la información y por esa vía mostrar evidencia de mayor credibilidad sobre los resultados de la investigación.

El investigador en el uso de la observación participante debe ser muy cuidadoso en cuanto el registro sistemático de los datos e información que va generando en el proceso de indagación en el grupo. Un recurso clásico en estos casos es el uso de las llamadas notas de campo, las cuales permiten al investigador llevar un registro pormenorizado tanto del escenario en sí, como de la vida del grupo, sus actividades y de la interacción social entre sus miembros. Al respecto, Taylor y Bogdan (1996) han señalado que:

Como método de investigación analítico, la observación participante depende del registro de notas de campo completas, precisas y detalladas. Se deben tomar notas después de cada observación y también después de contactos más ocasionales con los informantes, como por ejemplo encuentros casuales y conversaciones telefónicas [...] También debe tomarse notas durante la etapa previa al trabajo de campo (p. 74).

Estos mismos autores consideran que en estas notas de campo se deben registrar no sólo lo que el investigador observa en el escenario sino también los comentarios suyos acerca de sus propios sentimientos, interpretaciones, intuiciones, preconcepciones, dudas y áreas futuras de indagación. Además de las notas de campo, hay autores que recomiendan también el uso de dispositivos mecánicos o electrónicos para el registro de los datos e información producto del proceso de observación, tales como grabadores,

cámaras de video y/o cámaras fotográficas. Sin embargo, sobre este último aspecto, las opiniones de los autores están divididas en cuanto a la conveniencia o no del uso de estos dispositivos. Por ejemplo, Douglas (1976) sostiene que este tipo de recurso tecnológico atrae innecesariamente la atención del observador e interrumpen el flujo natural de los acontecimientos y conversaciones en el escenario.

Finalmente, algunas consideraciones sobre la ética que debe mantener el investigador durante el desarrollo del trabajo de campo. Al entrar al escenario y previo el *rapport* inicial establecidos con los miembros del grupo a observar, se produce un cierto nivel de confianza que permite a los actores sociales sentirse con libertad para actuar normalmente frente al investigador, lo cual le facilitara el tener acceso a información de muy variado tipo, incluso de naturaleza confidencial. Por tal motivo, el observador está en la obligación moral y profesional de proteger su fuente de información y no arriesgar o comprometer a los informantes bajo ninguna circunstancia.

**Salida del escenario.** Cuando decidimos hacer una investigación con base en la observación participante, como técnica de investigación cualitativa, tenemos en mente un propósito definido que se expresa en varias preguntas que intentamos responder o en objetivos que deseamos alcanzar mediante el proceso de indagación. Iniciamos el proceso de observación y obtenemos datos e información para satisfacer nuestros intereses investigativos. Ahora bien, en qué momento parar este proceso?; es decir, cuándo salir del escenario. En realidad, no existe un criterio fijo al respecto. Por lo general, se piensa que cuando se ha obtenido suficiente datos e información que permiten responder las preguntas o evidenciar el logro de los objetivos que se habían previsto inicialmente es el tiempo de iniciar la retirada del campo. Glaser y Strauss (1967) incorporaron en los

procedimientos del método comparativo constante, el concepto de *saturación teórica de las categorías*, lo cual se produce cuando la nueva información obtenida en el proceso de investigación se hace redundante; es decir, no agrega valor a la información previa. Ellos consideran que éste debe ser el momento de emprender la retirada y dar por concluida la recolección de información.

***Análisis de datos e información.*** En todo proceso de investigación es mandatorio cumplir con la fase de análisis una vez que se ha concluido el proceso de obtención de datos e información durante el trabajo de campo. En el caso de la investigación cualitativa esta fase analítica se puede atender durante el desarrollo mismo del proceso de obtención de datos, como lo plantea el método comparativo constante (Glaser y Strauss, 1967) o al final de dicho proceso, como lo hacen los etnógrafos. Sin embargo, no daremos aquí ninguna información específica al respecto y remitiremos al lector al capítulo 11 de esta obra en el que se tratará detalladamente el tema del análisis de datos e información en este tipo de investigación.

## La Observación No-Participante

En la observación no-participante, el investigador entra al escenario, observa los hechos desde alguna posición privilegiada -sin interferir con el hecho o situación observada-, registra las acciones, interacciones y transacciones que se dan entre los actores de la situación, pero sin tomar parte directa en los procesos que observa. Es decir, en la utilización de esta técnica se cumple un proceso análogo al que se utiliza con la observación participante, pero con la diferencia de que en la primera el investigador juega un rol activo en la dinámica del hecho que estudia, mientras que en la segunda esto no ocurre.

Esta técnica puede ser de dos tipos: directa e indirecta. En la primera, como ya se ha señalado, el investigador observa y registra los hechos, pero no forma parte del proceso; mientras que en la segunda los hechos se estudian con base en datos ex-post-facto, tales como anuarios estadísticos, censos o fuentes documentales (archivos, prensa y otros).

El uso de la observación no-participante directa tiene una larga historia en la investigación educativa, particularmente en lo que se refiere al estudio de las interacciones profesor-alumno que se producen en el salón de clase. Este tipo de estudios constituye un aspecto importante a la hora de realizar un análisis global del sistema educativo. Es por ello por lo que desde los años 30 se vienen realizando trabajos que reflejan la situación en las aulas y aportan ideas para su mejora (Buxarrais, 1989). Estos estudios han sido realizados tanto desde la perspectiva epistémica positivista como desde la óptica del paradigma interpretativo-naturalista.

Los estudios realizados bajo el paradigma positivista estuvieron enfocados al análisis de la conducta verbal y no-verbal de docente-alumnos y

sus interacciones. Los proponentes de este tipo de estudio parten del hecho de que el éxito del maestro puede ser juzgado a través del grado de efectividad de su enseñanza que puede ser objetivamente evaluada mediante su conducta o interacción en el aula. Así, el análisis sistemático y objetivo de la interacción en el aula puede proveer una evaluación confiable de lo que sucede dentro del salón de clase en términos de enseñanza y aprendizaje. Con este propósito se desarrollaron instrumentos de observación en el aula basados en sistemas de categorías. Esta forma de operacionalización se fundamenta en la conceptualización del llamado paradigma proceso-producto de investigación educativa, de orientación conductista, interesado en establecer relaciones entre las variables relativas al proceso de la enseñanza (comportamiento del profesor y de los alumnos y sus interacciones) y las variables que hacen referencia al producto de la enseñanza, entendido como el nivel de logro de los objetivos educativos por parte de los alumnos.

Estos instrumentos de observación en el aula basados en el uso de los sistemas de categorías, una vez aplicados eran sometidos a un proceso de codificación de los comportamientos verbales y no-verbales del profesor y de los alumnos procediéndose posteriormente a analizar los datos de observación así organizados.

Un clásico y relevante estudio en esta línea fue el desarrollo por Flanders (1973), quien propuso el modelo FIAC (Flanders Interaction Analysis Categories) enfocado hacia el estudio verbal de la interacción docente-alumnos a través de un instrumento de registro –rejilla- de 10 categorías de observación, las cuales estaban agrupadas de la siguiente manera (a) conducta de influencia indirecta del docente, tales como: acepta sentimientos, premia o estimula, acepta ideas, formula preguntas; (b) conducta de influencia directa del docente, por ejemplo, dicta la clase, da

instrucciones y formula criticas; y (c) conducta del alumno, tales como: inicia la conversación, responde a las preguntas, se queda en silencio.

La aplicación de este sistema implicaba la utilización de, al menos, dos observadores entrenados, ubicados en lugares distintos dentro del aula de clase. Cada observador codificaba minuciosamente las conductas pertinentes, tabulándose los símbolos del código en un esquema de presentación que ponía en claro las comparaciones deseadas. Todo el proceso tenía que efectuarse al mismo ritmo que el comportamiento espontáneo observado. La velocidad del registro solía estabilizarse al ritmo de una anotación entre tres y seis segundos (Buxarrais, 1989).

La tarea del observador consiste en codificar los eventos que se suceden en la clase en la categoría correspondiente del sistema, ya sea atendiendo a una unidad de tipo temporal -por ejemplo, registrando cada tres segundos los comportamientos observados-, ya sea atendiendo a una unidad natural/conductual -por ejemplo, registrando todas las apariciones de un determinado comportamiento (Coll y Sole, 1990). Posteriormente, a finales años 70 y comienzo de la década de los 80, fueron publicadas otras investigaciones sobre el estudio de la conducta de docentes y alumnos en el aula de clase. Así, se estudiaron, bajo observación no-participante, temas como: el comportamiento del docente y sus efectos (Brophy, 1979); el tiempo que los alumnos dedican a las tareas (Rosenshiney Berliner, 1980); organización y gerencia de aula (Doyle, 1992), entre otros.

Los estudios desarrollados bajo la orientación del paradigma interpretativo-naturalista, provienen de varias fuentes teóricas como son la psicología cognitiva de Piaget, el constructivismo socio-cultural de Vigotsky y la sociolingüística. En el primer caso, los estudios están orientados a estudiar la interacción estudiante-contenido en su proceso de auto

construcción del aprendizaje como individuo. Mientras que en el segundo, los estudios se dirigen más a la indagación sobre la construcción social del aprendizaje a partir de la mediación del docente y de los compañeros más aventajados. Finalmente, los estudios realizados bajo la orientación de la sociolingüística se interesan por la función del lenguaje en las interacciones entre el profesor y los alumnos, entre pares actuando como soporte para la adquisición de otros tipos de conocimiento.

En resumen, se podría señalar que la *observación participante* es una técnica de investigación cualitativa que ha sido utilizada tradicionalmente con gran éxito por antropólogos, sociólogos y psicólogos sociales, como parte del método etnográfico, para estudiar grupos étnicos, grupos marginales en comunidades urbanas y grupos sociales con determinada problemática psicosocial (jóvenes con trastornos de conducta, drogadictos, y alcohólicos, entre otros). Sin embargo, esta técnica de investigación ha sido utilizada con mucho menos frecuencia en el contexto de la investigación cualitativa que se realiza en educación.

En el contexto educativo, esta técnica podría ser utilizada de dos maneras: (a) cuando el interés de indagación está centrado en la institución escolar como objeto de estudio, en cuyo caso un investigador externo podría, previa autorización de las autoridades o a instancias de éstas, realizar un trabajo de campo orientado, por ejemplo, a la indagación sobre la problemática socioeducativa de la comunidad interna, la cultura organizacional (valores, creencias, actitudes, comportamientos) o cualquier otra situación de interés; y (b) cuando el foco del estudio está centrado en el hecho educativo que ocurre en el aula de clase. En esta situación particular, el investigador es el propio docente quien mediante el uso de la técnica de la

investigación-acción cumple el doble rol de facilitador del contenido y de investigador-observador al mismo tiempo.

La *observación no-participante*, por el contrario, es una de las técnicas de recolección de datos e información más utilizada en la investigación cualitativa que se realiza en educación, en el contexto del aula de clase, ya que este escenario de indagación no está ordinariamente abierto al público externo por su potencial interferencia con el desarrollo del hecho educativo. Creemos que su uso se incrementará en el futuro dado el espacio que progresivamente ha ido ganando el paradigma naturalista-interpretativo como modelo de investigación en educación.

## La Entrevista en Profundidad

En general, la entrevista puede ser definida como una técnica de interacción social dialógica que se realiza entre dos personas con el propósito de obtener información sobre un hecho o situación. Implica un proceso de comunicación que involucra a un entrevistador, quien formula las preguntas o propone los temas de conversación, y a un entrevistado quien responde libremente de acuerdo con su conocimiento, experiencia vivida e información disponible. Bingham y Moore (1973) consideran que entre la persona que entrevista y la entrevistada existe una correspondencia mutua, y gran parte de la acción recíproca entre ambas consiste en ademanes, posturas y gestos, entre otras conductas no verbales. Incluso las palabras adquieren diversos significados según su entonación y contexto. Todos estos elementos de comunicación concurren al intercambio intencionado de conceptos que constituye la entrevista.

En el contexto de la indagación social, la entrevista es una técnica de obtención de información oral o escrita en relación con un hecho o fenómeno de interés para el investigador. Puede ser de dos tipos: estructurada o abierta - también llamada cualitativa o en profundidad-. En la entrevista estructurada, el investigador intenta ser objetivo en el proceso de obtener la información requerida; para ello dispone de un conjunto de preguntas cuidadosamente planificadas y validadas, las cuales administra a un grupo de sujetos mediante el uso de procedimientos estandarizados.

La entrevista en profundidad, explora el mundo subjetivo del informante. La subjetividad en este contexto hace referencia al modo de pensar o sentir del individuo, y no al objeto en sí mismo; la subjetividad

desde el punto de vista filosófico hace referencia a las interpretaciones que se realizan sobre cualquier aspecto de la experiencia vivida, y es por esto que son sólo accesibles para la persona que vive dicha experiencia, ya que para cada individuo puede ser diferente, aunque se trate del mismo hecho o fenómeno.

A diferencia de la primera, la entrevista en profundidad tiende a ser flexible, dinámica, no estructurada, no estandarizada y está dirigida a indagar sobre el mundo interior de los individuos y a comprender el significado que les asignan a las experiencias vividas. En el contexto fenomenológico, profundizar quiere decir escudriñar, mirar a través de los ojos de la persona que tiene la vivencia, emitir opiniones y concluir sobre las experiencias del otro. En definitiva, lo importante y significativo en este tipo de entrevista cualitativa es describir acontecimientos y dimensiones subjetivas de las personas tales como: pensamientos, creencias, valores, actitudes, emociones y sentimientos. En este sentido, la entrevista profundiza cuando pasa del simple intercambio verbal entre el entrevistador y el informante al plano de intentar comprender la propia visión del mundo, las perspectivas y experiencias de las personas que son entrevistadas.

Taylor y Bogdan (1996) definen este tipo de entrevista como “reiterados encuentros cara a cara entre el investigador y los informantes, encuentros estos dirigidos hacia la comprensión de la perspectiva que tienen los informantes respecto a sus vidas, experiencias o situaciones, tal como las expresan con sus propias palabras”(p. 101). Por lo tanto, no se trata simplemente de obtener información a partir de la pregunta que formula el investigador para comprender o explicar un hecho social, sino más bien de descubrir el mundo de significados que subyacen en las verbalizaciones que hacen los informantes cuando relatan las experiencias vividas.

El modelo típico de una entrevista en profundidad es la conversación entre dos personas iguales, donde el investigador es el propio instrumento de investigación y no un técnico que hace preguntas y registra las respuestas con base en un cuestionario estandarizado. Algunos autores consideran que la entrevista nace de una ignorancia consciente por parte del entrevistador quien, lejos de suponer que conoce, a través de su comportamiento exterior, el sentido que los individuos dan a sus actos, se compromete a preguntárselo a los interesados, de tal modo que éstos puedan expresarlo en sus propios términos y con la suficiente profundidad para captar toda la riqueza de su significado (Ruiz Olabuenaga, 1996).

Las entrevistas en profundidad pueden ser clasificadas atendiendo a diferentes criterios; por ejemplo, según su propósito, pueden ser de tres tipos, a saber: el primero, tiene como fin “escarbar” sobre los hechos más relevantes acontecidos en la vida de una persona, expresada con sus propias palabras. Es una historia de vida o entrevista autobiográfica. A través de este tipo de entrevista, el informante revela su vida interior, expresada en su esfuerzo, luchas, éxitos y fracasos en el transcurrir del tiempo. El segundo tipo se dirige al aprendizaje sobre acontecimientos y actividades que no se pueden observar directamente. En este caso los informantes no solo ofrecen información al investigador acerca de su modo de ver la realidad, sino que deben describir lo que sucede y el modo de ver de otras personas. El tercer tipo tiene por finalidad proporcionar un cuadro amplio de una gama de escenarios, situaciones o personas. Esta se utiliza para estudiar un grupo grande de personas en un tiempo relativamente breve (Taylor y Bogdan, 1996).

Por otra parte, las entrevistas en profundidad pueden ser clasificadas según que busquen analizar muchos temas de forma general o se ciñan a un

tema concreto. En este sentido pueden ser categorizadas como: (a) holísticas, aquellas que están interesadas en todos los aspectos y puntos de vista concernientes al entrevistado; y (b) enfocada, las que se concentran en el esclarecimiento de determinados aspectos de una situación o estímulo. También pueden ser clasificadas según el control que impone el entrevistador sobre el informante y pueden ser dirigidas y no dirigidas. En el primer caso toda la iniciativa la lleva el entrevistador y, por lo tanto, el entrevistado se debe acomodar a los ritmos, desarrollos y variaciones posibles del investigador; mientras que en el segundo, el entrevistador tiene un mínimo control sobre el desarrollo de la entrevista. En consecuencia, el entrevistado plantea el ritmo de la acción, las preguntas y los temas de los que se va a hablar, así como el orden de los mismos y que el entrevistador no puede variar.

Si cruzamos las dos categorías anteriores: grado de generalidad (entrevistas holísticas y enfocadas) y grado de control (entrevistas dirigidas y no-dirigidas) se tendrían cuatro tipos de entrevistas en profundidad, a saber: entrevistas holísticas dirigidas, holísticas no-dirigidas, enfocadas dirigidas y enfocadas no-dirigidas.

La entrevista en profundidad, puede ser utilizada cuando: (a) los objetivos de la investigación son relativamente claros y bien definidos; (b) los informantes no pueden ser abordados de otra manera; (c) el investigador no dispone del tiempo suficiente como para ir a un trabajo de campo que le permita observar directamente el fenómeno que le interesa estudiar; (d) la investigación depende de una amplia gama de escenarios o personas; y (e) el investigador quiere esclarecer la experiencia humana subjetiva. Sin embargo, la entrevista en profundidad tiene sus limitaciones, por ejemplo: (a) puede ser susceptible de producir engaños y/o exageraciones; (b) puede producir

distorsiones de los hechos, en el sentido de que los informantes pudieran hacer o expresar cosas diferentes en distintas situaciones; y (c) el entrevistador desconoce el escenario en que han ocurridos los hechos.

Las limitaciones anteriores no son un motivo para descartar esta valiosa técnica de obtener información en la investigación cualitativa, como lo es la entrevista en profundidad. Pero es evidente que el investigador debe estar consciente de ellas para buscar los mecanismos apropiados que permitan neutralizar su posible efecto en la credibilidad de los resultados que se deriva de la misma. Por ejemplo, es importante que el investigador con su actitud contribuya a crear deliberadamente un clima de confianza que permita que el informante se abra sinceramente a la conversación. Por otra parte, el investigador puede indagar más sobre aquellos relatos que le parezcan contradictorios hasta precisar los hechos. Asimismo, cuando sea posible el investigador, si tiene dudas sobre un tipo de información puede buscar fuentes alternas de verificación, por ejemplo con familiares y/o amigos del informante, si fuera necesario.

## Momentos de la Entrevista

### Preparación

La preparación de la entrevista en profundidad requiere, por lo menos, prestar atención a dos aspectos relacionados con los futuros informantes, como son: (a) decidir quiénes serán y dónde encontrar los informantes clave; y (b) cómo contactar dichos informantes. A diferencia de la técnica de la entrevista estructurada, la entrevista en profundidad no requiere de una muestra estadística diseñada previamente con un número de informantes fijo y preseleccionados aleatoriamente para garantizar la representatividad de la muestra; se trata más bien de un diseño flexible en que apenas se definen algunos criterios preliminares para decidir a qué sujetos entrevistar de acuerdo con el tema, los objetivos del estudio y la intencionalidad del investigador.

En realidad, es difícil determinar a cuántas personas se debe entrevistar en un estudio cualitativo. Algunos investigadores tratan de entrevistar al mayor número posible de personas familiarizadas con un tema, hecho o fenómeno bajo estudio. Sin embargo, hay que tomar en cuenta que en la investigación cualitativa la validez de los resultados no está directamente relacionada con la cantidad de sujetos entrevistados, sino más bien con la calidad y variación de la información que aportan los entrevistados en relación con el tema.

Por otra parte, hay que considerar que las entrevistas cualitativas generan un volumen alto de información por lo cual se debe ser precavido para decidir el número de sujetos a entrevistar. En todo caso, sólo la dinámica de la investigación y los análisis de la información que se vayan realizando

sobre la marcha serán los que determinarán cuándo se ha obtenido la información necesaria y suficiente para decidir cuándo parar en la recolección de información. Sólo en ese momento se sabrá cuantos sujetos fueron requeridos por el estudio. Es decir, mientras que en la investigación cuantitativa, decidir sobre el tamaño de la muestra es un requisito *a priori*, en la investigación cualitativa este es un dato que se obtiene *a posteriori*.

En el método comparativo constante (Glaser y Strauss, 1967; Strauss y Corbin, 2002 ) se utiliza la estrategia del *muestreo teórico* según el cual el investigador aborda el primer informante, obtiene la información requerida, en primera instancia, regresa a casa o a la oficina, analiza la información (codificación, categorización, conceptualización) y decide cual es el próximo paso; por ejemplo, abordar un nuevo informante o regresar al informante inicial para profundizar, ampliar, aclarar o complementar algún aspecto de la entrevista anterior; repite el procedimiento de análisis, compara con los resultados previos y de nuevo decide cual será el próximo informante a entrevistar, que puede ser uno nuevo o volver al primero para una tercera entrevista y así sucesivamente hasta considerar que los nuevos informantes entrevistados no aportan ya información nueva. Se dice entonces que las categorías analíticas han sido saturadas y se procede a cerrar el proceso de recolección de información. Es en ese momento cuando se sabe cuántos informantes fueron utilizados en el estudio.

En este proceso de preparación de la entrevista en profundidad, Dolbeare y Schuman (citados por Seidman, 1998) proponen planificar por lo menos tres entrevistas con cada informante, las cuales se pueden realizar en días diferentes o el mismo día, dependiendo de la disponibilidad de tiempo del investigador. En la primera entrevista se establece el contexto y la experiencia del entrevistado. La segunda permite reconstruir los detalles de

su vida dentro del contexto en el cual esto ocurre, es decir, de su ahora; mientras que la tercera entrevista anima al entrevistado a reflexionar sobre el significado de sus experiencias.

En cuanto al segundo aspecto relacionado con el contacto con los informantes, se podrían utilizar cualquiera de los medios de información existentes dependiendo de las posibilidades de acceso que tenga el investigador (teléfono, correo, comunicación escrita, entre otras). En un estudio doctoral reciente, realizado bajo mi dirección, sobre las “Atribuciones de éxito y fracaso en las iniciativas institucionales de e-Learning” Rodríguez (2012) contactó los informantes (profesores universitarios con experiencia en educación virtual) mediante una comunicación escrita (ver ejemplo en página siguiente). En la misma se explicaba, entre otras cosas, el objetivo del estudio, su importancia institucional y se les invitaba a participar como informantes tomando en cuenta la experiencia en relación con la situación que se deseaba estudiar. Luego de enviada la comunicación, se le hizo un seguimiento telefónico a la respuesta hasta acordar la primera cita con el informante.

En el primer contacto con el entrevistado, es muy importante que el investigador se disponga a promover un clima psico-afectivo apropiado, que genere confianza y motive al informante para participar en el estudio. En otras palabras, en esta etapa es necesario que el investigador establezca un buen *rapport* a objeto de lograr la cooperación del entrevistado. Este primer contacto puede ser en forma personal, telefónica o de manera escrita. Algunos *tips* a ser considerados durante el desarrollo de la entrevista son los siguientes:

- Presentarse al informante aunque éste ya hubiese recibido la invitación

del investigador para participar en el estudio.

- Saber escuchar, mostrando atención y empatía a lo que el informante expone.
- Imitar sutilmente el lenguaje no verbal del entrevistado; por ejemplo, si éste cruzara los brazos o las piernas, el entrevistador podría hacer lo mismo unos segundos después, sin hacerlo de forma drástica.
- Mantener contacto visual, prestar atención al lenguaje no verbal (el movimiento de las manos, los gestos faciales, la postura corporal, el tono de voz) y mantener un porte relajado para que estas expresiones sean manifiestas.

En el primer contacto es importante también hacer algunas precisiones en cuanto a las intenciones del investigador con el trabajo a realizar, garantizar el anonimato del informante, el rol del entrevistado como validador de la información a ser publicada en el estudio y la logística para el desarrollo del trabajo (por ejemplo, horario y lugar de las entrevistas). Algunos autores recomiendan que al inicio de la primera entrevista, el investigador marque el tono de la relación con el informante, ser poco directivo y aparecer como alguien que desea aprender de la experiencia del entrevistado (Taylor y Bogdan, 1996; y Coles, 1971).

### *Ejemplo de Comunicación Escrita*

*Barquisimeto, 15 de abril de 2012*

*Ciudadano*

*Profesor Pedro González*

*Presente*

*Estimado Profesor:*

*Tengo el agrado de dirigirme a usted en la ocasión de saludarle, muy cordialmente, al tiempo que le expreso el planteamiento siguiente: soy la Prof. Haedy Rodríguez, estudiante del Programa Interinstitucional de Doctorado en Educación (PIDE) patrocinado por las universidades UCLA, UPEL y UNEXPO. Estoy actualmente iniciando mi proyecto de tesis, el cual versa sobre las “Atribuciones de Éxito y Fracaso en las Iniciativas Institucionales de E-Learning”. El escenario en el que aspiro desarrollar mi trabajo de investigación se relaciona con la experiencia que sobre Educación a Distancia ha tenido la UCLA, como institución pionera en este campo en la región.*

*Como base en lo anterior, he pensado en usted como uno de mis posibles informantes clave a objeto de obtener suficiente información sobre la percepción que tiene sobre esta experiencia educativa. En tal sentido, me gustaría conocer su disposición para participar en este proyecto, dándome la oportunidad de entrevistarle en su oficina en la fecha y hora que usted estime más conveniente.*

*Para la realización de este trabajo he entrado en contacto con el Dr. Arsenio Perez, Director de Educación a Distancia de esa Universidad, quien me ha autorizado para entrevistar a los profesores seleccionados en la muestra del estudio.*

*Sería de gran importancia y utilidad para mi trabajo de investigación, poder contar con su experiencia y testimonio, al permitirme indagar directamente, desde las vivencias de los propios actores, cómo han vivido este tipo de práctica educativa en esa universidad.*

*Como es normal en estos casos, la información obtenida de parte de los sujetos de la muestra será confidencial. En consecuencia, en la publicación de las ideas aportadas durante la entrevista se protegerá siempre la fuente. Además, la información obtenida será validada con el propio informante antes de su publicación a objeto de evitar cualquier error o sesgo en la interpretación de la misma.*

*El presente proyecto ha sido aprobado por mi Comité Académico, el cual está integrado por los doctores: Carlos Ruiz Bolívar (Tutor), Alirio Dávila y Luis Gómez, de quienes he recibido todo el apoyo y respaldo académico para la ejecución del mismo.*

*Segura de que podré contar con su desinteresada colaboración como informante en mi proyecto de tesis, me suscribo de usted expresándole mi sentimiento de la más alta estima y consideración.*

*Atentamente,*

Sobre este contacto inicial en la entrevista, Taylor y Bogdan plantean que:

El sello autenticador de las entrevistas cualitativas en profundidad es el aprendizaje sobre lo que es importante en la mente de los informantes: sus significados, perspectivas y definiciones; el modo en que *ellos* ven, clasifican y experimentan el mundo. Es presumible que los investigadores quieran formular algunas preguntas generales antes de iniciar el trabajo. Pero deben ser cuidadosos para no forzar su programa demasiado tempranamente. Al plantear de entrada preguntas directivas, el investigador crea una tendencia mental en los informantes acerca de aquello sobre lo que es importante hablar; esa predisposición inducida puede hacer difícil, si no imposible, llegar a conocer el modo en que realmente ellos ven las cosas (pp. 114-115).

Finalmente, un aspecto a tomar en cuenta en la preparación de la sesión de interacción con el informante se refiere a la guía de la entrevista. Como ya se ha mencionado al inicio de este capítulo, la entrevista en profundidad se asemeja mucho a la manera como se desarrolla una conversación entre dos personas; sin embargo, a diferencia de ésta, la entrevista se realiza con un propósito definido, cual es el de obtener un tipo de información determinado en función de unos objetivos de investigación. Por lo tanto, requiere de un mínimo de control del tipo de información que el entrevistador le pedirá al informante. Es en este contexto en el que surge la guía de la entrevista como un recurso flexible valioso que nos permite mantener el proceso de investigación en el marco de su intencionalidad.

Pero, es bueno aclarar, de una vez por todas, que la guía de la entrevista no es una lista de preguntas estructuradas y específicas sobre determinados temas a ser planteadas al entrevistado para obtener respuestas concretas, sino que se trata más bien de una lista de posibles temas a ser considerados durante el desarrollo de la entrevista. Dicha lista deberá ser manejada discrecionalmente por el investigador de acuerdo con la dinámica y ritmo que se siga en la entrevista y podría ser ampliada con nuevas preguntas o temas en la medida que el proceso de investigación vaya avanzando.

### **Desarrollo de la Entrevista**

Durante el desarrollo de la entrevista es importante que el investigador tenga presente que las palabras que expresa el informante no tienen un significado unívoco para todo el mundo, sino todo lo contrario mucho términos utilizados por el entrevistado pueden ser comprendidos e interpretados de diferentes maneras por la audiencia. Por ello, el entrevistador no debe dar por supuesto que entiende exactamente todo lo que el informante le quiere decir. De allí la necesidad de que el investigador solicite al informante, de vez en cuando,

aclaraciones y precisiones de aquellas expresiones u opiniones que parecen poco claras durante el desarrollo de su discurso.

Algunas estrategias para inducir claridad y precisión en lo dicho por el informante pueden adoptar la forma de parafraseo de lo dicho previamente por el informante. Asimismo, se pueden formular preguntas con el mismo propósito, por ejemplo, ¿Qué quiso decir cuando señala que los estudiantes son unos ineptos...? o ¿Cómo se sintió usted realmente ese día cuando el estudiante le grito en el aula de clase?, ¿a qué atribuye usted este tipo de comportamiento?, ¿podría usted poner un ejemplo de lo que quiere decir? Estas y otras preguntas ingeniosa y oportunamente formuladas podrían ser gran ayuda en el proceso de depurar la información obtenida durante el desarrollo de la entrevista.

Un aspecto a ser considerado durante el desarrollo de la entrevista es el registro de la información que va emergiendo de la comunicación investigador-informante. Al respecto los investigadores tienen diferentes criterios. Por ejemplo, algunos consideran que se debe confiar en la memoria y tomar notas sólo de algunas expresiones o palabras clave que desee retomar posteriormente. Se aduce que el rol principal del investigador es proponer los temas de conversación, controlar el clima psico-afectivo de la interacción comunicacional y prestar atención a lo que el informador dice y la manera como la hace (comportamiento no-verbal).

Otros investigadores, por el contrario, admiten la posibilidad de que se puedan utilizar artefactos electrónicos, tales como el grabador o la video-grabadora, aunque reconocen que al inicio de la entrevista este tipo de artefacto puede ser un distractor que inhiba el comportamiento verbal del informante. En todo caso, se recomienda que para utilizar grabador durante la entrevista, el investigador debe solicitar autorización al entrevistado.

Asimismo, se sugiere que su colocación en el sitio de la entrevista deba ser discreta o poco visible a objeto de controlar el posible efecto inhibitor de dichos equipos.

Hay autores quienes recomiendan a los investigadores llevar un diario, como el propósito de ir registrando día a día, después de cada entrevista, algunos incidentes críticos que pudieran ser de gran utilidad para volver a ellos posteriormente, o para ayudar en la interpretación posterior de la información. También el diario sirve para que el investigador pueda ir registrando algunos comentarios que le vayan surgiendo en la medida que transcurre el proceso de investigación. Al respecto, Taylor y Bogdan (1996) consideran que los siguientes son ejemplos del tipo de comentarios que deben incluirse en el diario:

- “Por la cara que ponía, creo que ella ironizaba al hablar sobre su madre. Pero no parecía querer decir nada realmente negativo sobre ella”.
- “Es la tercera vez que plantea el tema. Debe ser importante para ella. Tengo que estudiar esto en el futuro”.
- “De algún modo los dos estábamos aburridos esta noche. Sólo queríamos que la entrevista terminara. Quizás esto se debió al tema o tal vez ambos estábamos cansados hoy”.
- “Creo que fui demasiado agresivo esta noche. Me pregunto si dijo esas cosas sólo para que yo no lo apremiara. Debo tenerlo presente cuando repase la conversación”.

### **Cierre de la Entrevista**

Una recomendación final es indicada al momento de concluir la entrevista.

En ese instante, el investigador podría hacer un resumen de los aspectos más relevantes tratados durante la sesión y solicitar al informante su confirmación y ofrecerle la posibilidad de agregar cualquier información adicional, si lo desea. Asimismo, se recomienda que el investigador exprese su agradecimiento al entrevistado por su tiempo y por los aportes hechos a la investigación. También, se debe dejar acordado la fecha y hora de la próxima sesión, si fuera necesario. Igualmente, pudiera quedar pendiente una posible llamada posterior para concertar una nueva entrevista si hiciera falta.

Una etapa final, en el proceso de la investigación cualitativa basada en la técnica de la entrevista en profundidad, consiste en el análisis de la información obtenida, lo cual permitirá darle organización y sentido a la misma en el contexto del cumplimiento de los objetivos de la investigación. Para tal fin, existen diferentes técnicas de análisis, así como procedimientos manuales y electrónicos para facilitar esta ardua tarea. Sin embargo, este aspecto no será tratado en esta sección y se dejara para ser desarrollado en el capítulo 11 de esta obra, el cual será dedicado específicamente al tema del análisis de la información en la investigación cualitativa.

En el contexto educativo, la entrevista en profundidad es una de las técnicas de investigación cualitativa más utilizada, particularmente en el nivel de la educación superior. Recientemente mis estudiantes de doctorado la han empleado en sus trabajos de tesis. Así, Rodríguez (2012) la utilizó para estudiar las atribuciones causales de éxito y fracaso que hacen los decisores y actores sobre el proceso educativo virtual. El escenario utilizado para la indagación fue la Dirección de Educación a Distancia de la Universidad Centroccidental Lisandro Alvarado (UCLA), ubicada en la ciudad de Barquisimeto (Venezuela). También Briceño (2012) utilizó la entrevista en profundidad en su investigación sobre la construcción del éxito académico en

estudiantes de ingeniería en la Universidad Politécnica Experimental “Antonio José de Sucre” en la ciudad de Barquisimeto (Venezuela). Asimismo, Mujica (2012) en su proyecto de tesis doctoral previó utilizar la entrevista en profundidad como una de las técnicas de indagación en su estudio sobre la cultura organizacional universitaria y el uso de las TIC.

# Grupo Focal y Observación Flotante

## Grupo Focal

### Introducción

Como se ha señalado ampliamente en la Parte II de esta obra, la investigación cualitativa se interesa por la vida interior de las personas. Por ello indaga acerca de cómo los individuos viven sus experiencias de la vida diaria, sus sentimientos y los significados que le asignan, con el propósito de comprender, describir e interpretar la información que se obtiene a través de medios flexibles, no intrusivos que recogen información básicamente de naturaleza verbal. Una de esas técnicas de investigación es el grupo focal, también llamado grupo de enfoque o “focus group” (en inglés).

La literatura reporta diferentes definiciones sobre la técnica del grupo focal, las cuales coinciden en aspectos tales como los siguientes: (a) es una discusión informal en un grupo de individuos seleccionados para intercambiar ideas sobre un tema específico; (b) son reuniones semiestructuradas, coordinadas por un moderador entrenado; (c) su propósito es el de recolectar experiencias y creencias personales relacionadas con el tema designado; y (d) para comprenderlo en profundidad desde la perspectiva de los participantes del grupo (ver, Liamputtong, 2010; Kitzinger 1994; Wilkinson 2004; Morgan (2002); Kruger y Casey, 2000; entre otros). Con base en la revisión anterior, el grupo focal puede ser definido en los términos siguientes: es una técnica de investigación cualitativa que permite obtener datos e información a partir de una discusión grupal informal (7 a 11 personas), bajo la coordinación de un moderador, sobre un tema en particular, con el propósito de comprender los significados e interpretaciones de los participantes expresados en sus propias palabras.

Dicho de otra manera, a través del grupo focal, el investigador busca conocer las experiencias subjetivas de los miembros del grupo con respecto al tema objeto de la discusión. En consecuencia, el investigador-moderador prestara atención a la opinión de los sujetos (conducta verbal) y la manera como la expresan (conducta no-verbal) y tanto a la información fáctica como a las creencias, sentimientos, emociones y actitudes de los individuos hacia el problema o tema de discusión. Al respecto, Kruger y Casey (2000) agregan que en el grupo focal los participantes influncian y son influenciados, mientras los investigadores juegan diferentes roles, incluyendo el de moderador, oyente, observador y eventualmente el de analista inductivo.

Como técnica de investigación, el grupo focal se basa en dos supuestos básicos: primero, que los miembros del grupo constituyen una rica fuente de información sobre el tema de estudio; y segundo, que las respuestas individuales y colectivas estimuladas por la situación del grupo focal generará un material informativo que difiere del obtenido a través de otras técnicas de indagación ante un problema similar (Glitz, 1998).

Los elementos clave que contribuyen a que el grupo focal sea un instrumento efectivo para el científico social son el nivel de sinergia, la expansión de las ideas, la estimulación y la espontaneidad que una dinámica de grupo puede generar (Catterall & Maclaran, 1997), todo lo cual hará que la acción del grupo como tal se concentre en dirección al objetivo de la discusión, lo que permitirá lograr un efecto tipo “bola de nieve” en el que el comentario de alguien estimule nuevas ideas en otro participante, las cuales podrían tener un efecto generativo en otros miembros del grupo y así sucesivamente.

El grupo focal ha evidenciado tener algunas ventajas como técnica de investigación social, en comparación con otras técnicas cualitativas

(entrevistas individuales en profundidad y observación participante) y cuantitativas (cuestionario, escalas de estimación e inventarios); algunas de ellas son:

- Permite conocer las opiniones individuales y grupales, favorables y desfavorables de los participantes con respecto al tema objeto de discusión.
- Se obtiene información detallada tanto de tipo fáctico como del dominio psicoactivo del sujeto, referido a los sentimientos, actitudes, emociones, valores y creencias de los participantes con respecto al tema.
- La flexibilidad, espontaneidad y libertad de los participantes para expresar sus ideas sobre el tema en discusión permite obtener información valiosa y en profundidad que difícilmente se puede lograr con los instrumentos tradicionales de investigación en ciencias sociales (cuestionarios, escalas de estimación, inventarios).
- En comparación con otras técnicas de investigación cualitativa permite conocer un tema en profundidad y con credibilidad en un tiempo relativamente corto.
- Hace posible explorar temas nuevos para la formulación de hipótesis a ser comprobadas en estudios posteriores.
- Se generan ideas y conceptos nuevos, los cuales permiten ampliar la comprensión sobre el tema objeto de investigación.

No obstante lo anterior, esta técnica de investigación también es

portadora de algunas limitaciones, como son las siguientes:

- La calidad de los resultados es excesivamente dependiente del liderazgo y del entrenamiento previo del moderador.
- Sus resultados solo pueden ser interpretados en el contexto de los estudios de caso.
- Puede haber tendencia en algunos participantes a monopolizar la discusión y a imponer sus ideas, lo cual resulta peor cuando el moderador es inexperto.
- Se requiere de una logística de apoyo altamente eficiente (accesibilidad al local, confortabilidad de ambiente físico, compensación del tiempo de los participantes, refrigerios, material de apoyo).
- El análisis de los resultados puede resultar complejo y tedioso.

### **Antecedentes**

La técnica del grupo focal ha tenido una larga historia en la investigación en ciencias sociales. Entre los años 20 y 40 fue utilizada informalmente por diferentes investigadores sociales con el nombre de entrevista en grupo, como ha sido reportado en los trabajos de Montero (2009), Madriz (2003), y Frey y Montana (1993). Sin embargo, no fue sino hasta la II Guerra Mundial cuando se formalizó su uso como técnica de investigación social. Esta técnica fue utilizada desde la Universidad de Columbia para estudiar el impacto de los medios en la actitud de la gente hacia el involucramiento de los Estados Unidos en las operaciones bélicas de la Guerra (Merton & Kendall 1946, citado por Liamputtong, 2010).

Lazarsfeld y Melton invitaron a un grupo de sujetos para que oyeran un programa de radio que había sido creado para subir la moral de la gente en relación con el esfuerzo por la guerra (Merton, 1987). Originalmente se les pidió a los participantes pulsar un botón electrónico para indicar su respuesta en forma positiva o negativa el programa de radio. Sin embargo, la información obtenida no ayudó a los investigadores a descubrir por que los participantes respondieron en la forma que lo hicieron.

Ello puso en evidencia que este método no era suficiente para comprender la complejidad del punto de vista de los sujetos (Conradson, 2005). En subsecuentes estudios, los investigadores desarrollaron un enfoque alternativo para llevar a cabo este tipo de entrevista grupal. En esta propuesta se prestó mayor atención a los aspectos cualitativos, no-estructurados del punto de vista de los participantes, expresado con sus propias palabras. En lo sucesivo estos investigadores aunque seguían usando el grupo focal como una estrategia de investigación cualitativa, ellos lo hicieron con un propósito exploratorio como un medio para generar preguntas que pudieran ser usadas para desarrollar nuevas estrategias cuantitativas o para complementar los resultados en este tipo de investigaciones (Madriz, 2003).

Como se puede ver, el grupo focal nació en el contexto de la investigación académica, como una técnica de investigación sociológica. Sin embargo, pareciera que su mayor auge ha estado en el campo de la investigación de mercado y la publicidad, donde ha tenido un gran desarrollo. Después de los años 80 esta técnica se reinsertó en el ámbito académico y actualmente convive paralelamente en ambos. De acuerdo con Montero (2009), esta reincorporación de la técnica del grupo focal a la investigación científica se debió a dos factores: (a) la adopción de dicha técnica en el campo de la política y de las ciencias que la estudian; y (b) el amplio

desarrollo logrado por el paradigma naturalista-interactivo como marco onto-epistemico y metodológico para el estudio de los fenómenos sociales.

## Aplicaciones de la Técnica del Grupo Focal

De acuerdo con Williams y Katz (2000), la técnica del grupo focal puede ser utilizada cuando la investigación tiene algunos de los propósitos siguientes: (a) explorar sentimientos, percepciones, motivaciones, creencias, actitudes y valores; (b) comprender por qué los consumidores rechazan de una cierta manera un producto o el mensaje de una publicidad; (c) sugerir guiatra para el desarrollo de procesos; (d) explorar problemas para formular hipótesis cuando éstas no existen; (d) comprender, desde un punto de vista cualitativo, el mensaje que subyace detrás de los números; y (e) proveer de información de entrada acerca de problemas que deberían ser medidos usando técnicas cuantitativa.

En el caso de la educación, durante la última década la técnica del grupo focal ha sido utilizada en el contexto internacional en los niveles de la educación media y superior con diferentes propósitos. Sin embargo, es importante estar consciente que la finalidad del grupo focal es obtener información, para facilitar la toma de decisiones acerca del funcionamiento de una organización o programa educativo para guiar su acción futura, pero durante el proceso de desarrollo del mismo, como tal, no se toman decisiones, no se resuelven problemas, ni se completan tareas relacionadas con la institución y/o programa.

La información obtenida mediante esta técnica ha sido usada para: determinar necesidades de programas educativos, diseño de programas, evaluación piloto del currículo y sus productos, mejoramiento de programas, satisfacción de los usuarios con el programa, desarrollo y evaluación de políticas, evaluación de resultados, el desarrollo de instrumentos de

aprendizaje que atraerán las necesidades e intereses de los estudiantes, evaluación del conocimiento y/o actitudes de los estudiantes acerca de los problemas del currículo, y en la ampliación y complementación de los resultados de investigaciones educativas basadas en encuestas (Rennekamp y Nall, 2001).

### **Metodología del Grupo Focal**

La metodología que nos permite utilizar de manera eficiente la técnica del grupo focal, sigue la misma estrategia que cualquier otra con un propósito similar, cual es el de obtener información creíble y útil sobre un problema o tema de estudio. En este sentido, es necesario atender tres fases o momentos, a saber: la planificación, el proceso de ejecución y análisis de resultados. Estas fases se describen a continuación.

**Planificación.** Como se ha señalado anteriormente, una de las ventajas que tiene la técnica del grupo focal es que hace posible lograr información creíble y relevante para la comprensión del problema objeto de estudio. Sin embargo, esta cualidad no es el resultado de una acción más o menos improvisada por parte del investigador, sino que ello requiere de una cuidadosa planificación que permita prever los elementos, acciones y recursos necesarios para garantizar el logro del objetivo previsto con la reunión del grupo. A continuación se describen cada uno de los aspectos que muchos autores consideran se deben tomar en cuenta en el momento de planificar la utilización de dicha técnica de investigación.

- **Objetivo del grupo focal.** Todo grupo focal debe tener un propósito y este tiene que estar en relación directa con el objetivo de la investigación, para la cual se intenta recolectar los datos e información.
- **Selección de participantes.** El primer criterio que se debe tener en cuenta

para decidir quienes serán los participantes potenciales que participarán en la reunión-discusión de un grupo focal es el dominio del tema (experiencia vivida, conocimientos). En segundo lugar, estaría la heterogeneidad del grupo en cuanto a posibles opiniones diferenciales con respecto al tema a ser discutido. Existen diferentes criterios en cuanto al número de participantes a seleccionar, los cuales están comprendidos en un rango entre 5 y 12. En particular, me acojo al criterio propuesto por Montero (2009), quien recomienda y justifica por qué el tamaño del grupo debe estar entre 7 y 11 participantes.

- **Reclutamiento de participantes.** Para el reclutamiento, es recomendable hacerlo por la vía informal, primero, mediante un contacto personal, y luego formalizar la invitación mediante una comunicación escrita donde se precisen los detallares de la reunión-discusión (objetivo, justificación e importancia, lugar de la reunión, fecha y hora). Se debe precisar si se hará alguna recompensa a los participantes por su participación en la reunión (constancia de participación, diploma, dinero u otro estímulo).
- **Selección del moderador.** El funcionamiento apropiado de un grupo focal requiere de un buen moderador, el cual pudiera ser el propio investigador u otra persona debidamente entrenada para desempeñarse como tal. Un moderador eficiente debe tener la habilidad de saber escuchar, indagar, motivar y coordinar las interacciones entre los miembros del grupo. El moderador debe sentirse confiado y comfortable en el desempeño de su rol y generar confianza en el grupo.
- **Guión de preguntas.** Consiste en una lista de aspectos relacionados con el objetivo de la investigación que deben ser presentados al grupo para su

discusión. No se trata de formular preguntas directas a los miembros del grupo para obtener las respuestas respectivas, sino más bien de plantear temas, por parte del moderador, a objeto de que sean debatidos por el grupo. En este sentido, el moderador debe estar alerta para cambiar de tema cuando sienta que el punto en discusión ya ha sido suficientemente debatido (la discusión no agrega información nueva).

- **El lugar de reunión.** Es muy importante que el lugar de la reunión sea de fácil acceso para los participantes invitados. Asimismo, se sugiere que el salón de unión sea apropiado en términos de su funcionalidad y confortabilidad. Se deben evitar las interrupciones telefónicas.
- **Duración de la reunión.** Es necesario prever el tiempo que durará la reunión-discusión. Al respecto, no existe un criterio único, ya que ello depende de factores tales como la naturaleza y complejidad del tema y el nivel educativo de los miembros del grupo. En la literatura se encuentran recomendaciones de investigadores que van desde un mínimo de 45 minutos hasta un máximo de dos horas “según el entusiasmo, interés y productividad demostrados por sus integrantes así como su aceptación o proposición de alargarlo un poco mas” (Montero, 2009, p. 53).
- **La logística.** Es muy importante prever un refrigerio (café, jugos, pasta seca, cachitos, sándwiches, chocolates) para servir a los participantes al inicio, durante o al final de la reunión. Igualmente, se recomienda que se facilite a los participantes papel y lápiz para tomar notas si lo consideran necesario.
- **Normas de participación.** Se considera útil, a los fines de lograr la

mayor eficiencia en el trabajo grupal, que hayan unas normas que regulen el funcionamiento del mismo. Por ejemplo, el tiempo de participación por persona; este puede variar entre 2 y 3 minutos, con un máximo de 3 a 4 intervenciones por individuo. Las intervenciones deben estar referidas al tema, no se deben formar diálogos entre los participantes vecinos y se deben apagar los celulares.

- **Ejecución.** Una vez que se han tomado todas las previsiones del caso, se procede a instalar la reunión del grupo en el lugar y hora convenido. Durante esta fase, la actuación del moderador tendrá particular relevancia al inicio de la reunión, durante el desarrollo de la discusión y al cierre de la misma. A continuación se describen las principales actividades que se supone debe desarrollar el moderador en cada uno de estos momentos.
- **Inicio.** La reunión debe comenzar puntualmente y es recomendable que el moderador se encuentre en la sala previamente para chequear que todo esté en orden y a objeto de recibir a los participantes invitados. Una vez completado el grupo, el moderador deberá introducir la reunión, para lo cual le dará la bienvenida al grupo, expondrá el objetivo de la reunión, la duración de la misma y las normas de funcionamiento. A continuación se organizará una dinámica de presentación de los participantes. En este sentido, se les podría solicitar a los ellos que indiquen sus respectivos nombres y apellidos, ocupación y su relación con el tema a ser discutido. Para “romper el hielo” se propone realizar una breve dinámica de grupo. Por ejemplo, se le podría pedir a los participantes que respondieran la pregunta siguiente: *Qué harían ustedes con un camión de pétalos de*

*rosas rojas: indicar dos acciones positivas y dos negativas.* Finalizada la dinámica, se plantea el primer punto de la agenda y se inicia la discusión.

- **Desarrollo.** Una vez que se inicia la reunión, el moderador se mantendrá atento a la dinámica de la misma tratando de captar lo más posible el contenido de la discusión y las expresiones no verbales de los participantes. Es altamente recomendable disponer de una videograbadora o, en su defecto, de un grabador de audio a fin de registrar la información que se vaya generando en la reunión-discusión. Cuando cada punto de la agenda se haya agotado, el moderador deberá estar pendiente para proponer al grupo pasar al tema siguiente. Asimismo, deberá estar alerta para formular preguntas a los individuos o al grupo cuando lo estime necesario, ya sea para aclarar, ampliar, profundizar o complementar las ideas de la discusión.
- **Cierre.** Cuando se haya agotado la agenda o se hayan discutidos todos los temas previstos en el guión preestablecido, el moderador procederá a cerrar la reunión, para lo cual dará las gracias a los asistentes por su colaboración y se les informará sobre la entrega de la compensación (diploma, dinero, obsequio) en el caso de que estuviera previsto hacerlo.

### **Análisis de los Datos e Información**

Finalizada la sesión de la reunión-discusión, el moderador-investigador procederá a transcribir la información obtenida, prestando especial atención no solo al contenido (qué se dijo) sino también a la comunicación no-verbal (cómo se dijo) reflejadas en el tono de la voz, gesticulaciones y otras expresiones corporales. En un segundo momento, se procederá a establecer la unidad de análisis (palabra, frase, párrafo) para su posterior codificación y

categorización, como parte del proceso de análisis. En el capítulo 11 de esta obra se describen detalladamente los diferentes métodos de análisis tanto mediante el uso de un procedimiento manual como electrónico.

## La Observación Flotante en los Entornos Virtuales

### Contextualización del Escenario

#### **Contextualización del Escenario**

El impacto de las tecnologías de la información y la comunicación (TIC, Internet, computadores digitales, teléfonos celulares inteligentes, iPad o tablets, Web 2.0 y redes satelitales, ente otras) sobre la educación ha generado diversos modelos tecno-pedagógicos de aprendizaje (e-learning, b-learning, m-learning, u-learning) que están dejando atrás el viejo modelo instruccional tradicional. Estos nuevos modelos educativos se caracterizan porque: (a) el aprendizaje esta centrado en el estudiante; (b) los participantes forman comunidades de aprendizaje que funcionan como un mecanismo de cooperación y apoyo mutuo en el proceso de construcción social del conocimiento; (c) el diseño instruccional se planifica y gestiona desde una plataforma o sistema de gestión del aprendizaje; (d) el profesor juega el rol de tutor que orienta, da retroinformación y ofrece ayuda oportuna a los estudiantes para superar sus dificultades de aprendizaje; y (e) la comunicación entre los actores del proceso (docente-alumnos y alumnos-alumnos) se puede hacer de manera asincrónica o sincrónica creando, a su vez, un espacio singular y novedoso para la investigación educativa de tipo fenomenológico.

La fenomenología, como se señaló en la Introducción de esta obra, puede ser definida como el estudio de los fenómenos tal como son experimentados, vividos y percibidos por el ser humano (Martínez, 1989). En consecuencia, esta orientada a la comprensión del significado de las organizaciones y de los individuos que las integran a partir de la percepción

que el investigador tiene de la interacción entre ellos y los significados que le asignan a la experiencia vivida. Su propósito es describir y reflexionar sobre los fenómenos humanos estudiados. Se trata de un enfoque intersubjetivo en el que el investigador y el investigado, intercambian puntos de vista, de manera que el uno pueda estar en el lugar del otro y así entender e interpretar de manera más fidedigna la realidad presentada. Intenta reducir a su mínima expresión las influencias de teorías e ideas previas, no se orienta por hipótesis, y procura captar toda la realidad que se presenta de manera vivencial a nuestra conciencia, en el sentido planteado por Husserl (1962). En este enfoque el investigador utiliza diferentes técnicas y procedimientos para lograr tener acceso al mundo subjetivo del individuo; una de ellas es la entrevista en profundidad.

En la investigación fenomenológica que se realiza en los entornos virtuales, la información que expresa el mundo subjetivo del estudiante sobre la experiencia de aprendizaje vivida proviene, por una parte, de la interacción directa, en tiempo real, de los estudiantes con el profesor-tutor mediante la utilización de herramientas tales como: la videoconferencia y el chat, entre otros. Por otra parte, los datos se obtienen por vía asincrónica representada por la información textual producto de las acciones, interacciones y transacciones que ocurren entre los estudiantes en el proceso de construcción social del conocimiento. La información que se analiza queda registrada en la plataforma o sistema de gestión de aprendizaje, como ocurre en el caso de los foros de aprendizaje, foros sociales, talleres de solución de problemas, herramienta Wiki, Weblogs y los mensajes de correo electrónico, entre otros.

En el caso de las experiencias virtuales de aprendizaje, apoyadas en una plataforma tecnológica o sistema de gestión del aprendizaje (LMS, por sus siglas en Inglés), como ocurre con el e-learning, se genera una información

valiosa como resultado del desarrollo de actividades y tareas propias del proceso de interacción didáctica en la construcción social del aprendizaje, que constituye la base fundamental para el análisis y estudio de dichos procesos. Esta información fundamentalmente es el resultado de las acciones, interacciones y transacciones que se producen entre los estudiantes y sus pares, y entre éstos y el tutor, en actividades tales como foros (de aprendizaje, sociales, de tutoría), chat, trabajo en wiki, o en materiales en formato de videos o audio, entre otros. Este tipo de información que se genera como un subproducto del proceso de interacción virtual puede ser estudiado con diferentes propósitos y mediante el uso de diferentes enfoques tales como los de la antropología, la lingüística y la psicopedagogía.

### **Enfoque Antropológico**

El estudio de las interacciones sociales que se producen en el ciberespacio, a través de la Internet ha recibido una gran contribución desde el campo de la antropología, particularmente mediante el aporte del concepto de cultura y del método etnográfico. Antropólogos como Mayans (2002), Faura (2000) y otros han defendido la pertinencia de los estudios sociales sobre el ciberespacio por diferentes motivos: por tratarse de un espacio donde se producen relaciones sociales (aunque estén mediadas por un ordenador), porque tiene su propio lenguaje, sus propias reglas y dinámicas, por el cambio sociocultural que ha supuesto, por la presencia y repercusión que tiene actualmente; en definitiva, porque se trata de un espacio social y una fuente inagotable de cuestiones que son susceptibles de ser estudiadas y analizadas por la antropología y otras ciencias sociales.

En este contexto ha aparecido en la literatura especializada una nueva terminología que alude a la aplicación de los conceptos antropológicos a los entornos virtuales; así, han surgido términos tales como: ciberantropología,

como una rama de la antropología sociocultural que estudia los sistemas cibernéticos y la relación entre los humanos y las nuevas tecnologías (Downey, Dumit & Williams, 1995); la cibercultura, como el conjunto de técnicas, de maneras de ser y de hacer, de valores, de representaciones que están relacionadas con la extensión del ciberespacio (Faura 2000) y la etnografía virtual, la cual se encarga del estudio de los procesos socioculturales que ocurren en y a través del ciberespacio (Hine, 2004).

En su sentido tradicional, la etnografía es uno de los métodos clásicos de investigación socio-antropológica que tiene como propósito el estudio de las tradiciones culturales de grupos étnicos o pueblos primitivos, expresadas como un resultado analítico-descriptivo de las costumbres, comportamientos, conocimientos, creencias, prácticas sociales y religiosas. Modernamente, los estudios etnográficos se han extendido al estudio de comunidades, organizaciones y grupos sociales que comparten objetivos, intereses, valores y prácticas comunes, como podría ser el caso de una escuela, un aula de clase o una comunidad virtual de aprendizaje. En este último caso, este método recibe el nombre de etnografía virtual, también denominada etnografía del ciberespacio (Hakken, 1999), etnografía online o ciberetnografía (Escobar, 1994).

Hine (2000) considera que el espacio concreto en que se mueve una etnografía del ciberespacio no es acotable ni similar a los clásicos emplazamientos físicos que han ocupado los grupos sociales que han acostumbrado a estudiar los antropólogos. También Strathern (1996) ha ensalzado el valor de una etnografía que deje de fijarse en espacios acotados (emplazados, estáticos, fijos) y sepa prestar atención a un tipo de espacio hecho principalmente de vínculos y conexiones, un espacio-red.

Por otra parte, Hine (2004) ratifica su propuesta de una nueva etnografía

de Internet basada en el estudio de hechos mediáticos concretos, donde la red tiene un papel, por un lado, de instancia de conformación cultural y, por otro, de artefacto cultural construido sobre la comprensión y las expectativas de los internautas. Él afirma que:

El desarrollo y la expansión de las nuevas tecnologías y específicamente de la red, no depende de las propiedades intrínsecas de éstas sino fundamentalmente del uso y el sentido que le dan quienes participan en ella, ya sea como productores de espacios web o como consumidores de los mismos. Es desde esta interacción – negociación que la vida online se multiplica generando ricos intercambios. La metodología que permite este descubrimiento, no es otra que la etnografía virtual (p. 20).

De acuerdo con lo anterior, la investigación etnográfica se centraría, entre otras cosas, en los entornos sociales y las relaciones que sus informantes o usuarios, establecen a través de Internet. Se considera que el contexto ciberespacial es un campo de estudio donde es posible analizar la interacción de los individuos o informantes cibernautas y la creación de una denominada “cibercultura” en la Red (Téllez-Infantes, 2002).

En este contexto, Etchevers Goijberg (2005) considera que el papel del etnógrafo en Internet no sólo permite acercarse a una realidad para describirla de manera detallada, sino que requiere que el investigador se introduzca en esta interacción mediada con el propósito de reflexionar sobre el uso directo del medio y el contexto en el que se efectúa. Ella agrega, además, que la etnografía virtual permite al investigador utilizar su propia experiencia y combinarla con los elementos encontrados en el estudio a objeto de contrastarlos con las teorías existentes.

Un ejemplo que ilustra la investigación en educación, con el método de

la etnografía virtual, es el desarrollado por Montero (2012). Ella estudió las relaciones, interacciones y transacciones que se establecieron entre los actores sociales de una comunidad virtual de aprendizaje (CVA), en el contexto del curso Diseño, Construcción y Validación de Instrumentos de la maestría en Investigación Educativa de la Universidad Pedagógica Experimental Libertador, Instituto Pedagógico de Barquisimeto (Venezuela). Se utilizó una muestra integrada por 38 participantes correspondientes a dos secciones de clase; eran mujeres y hombres, educadores de profesión. La selección de los actores sociales fue de manera intencional, atendiendo primeramente a su participación activa en el curso. En tal sentido, se les invitó a formar parte de las diversas actividades y escenarios propios del mismo, los cuales se diseñaron y adaptaron a los espacios educativos virtuales. Esto permitió que ellos expresaran sus opiniones, ideas, sentimientos y construcciones generados en el transcurrir de la experiencia de aprendizaje. La metodología utilizada fue la etnografía virtual y como técnica de recolección de datos se empleó la observación flotante. El examen de la información se realizó con base en técnicas de análisis cualitativo, las cuales contemplan la búsqueda de unidades de interpretación, la codificación y la categorización. De esta manera emergieron categorías, subcategorías que fueron trianguladas entre las mismas y las teorías de base. Los hallazgos de esta investigación permitieron la construcción de una aproximación teórica que condujo al desarrollo de una cosmovisión con respecto al aprendizaje virtual en el cual se vislumbraron significados cognitivos, afectivos y sociales que fueron negociados y asumidos en el convivir de la CVA estudiada.

### **Enfoque Lingüístico**

Como ya hemos señalado, la información que se genera en las interacciones educativas en los entornos de aprendizaje virtuales es, en gran medida, de

naturaleza textual. Es decir, se trata de un tipo comunicación mediada por el computador que privilegia el discurso escrito. Por ello, tiene relevancia los aportes que la disciplina de la lingüística nos ofrece como opción para interpretar la información que se produce en las acciones, interacciones y transacciones que ocurren en la comunicación entre docente-alumnos y alumnos-alumnos en el proceso de construcción social del conocimiento. Estas interacciones normalmente ocurren mediante la discusión en foros de aprendizaje, el trabajo en equipos colaborativos, la interacción en los foros sociales, videoconferencias, participación en blogs y, el uso de herramientas como el correo electrónico y el chat.

La lingüística se conoce como la ciencia que estudia el lenguaje humano articulado en sus aspectos teóricos y prácticos. En el primer caso, se trata de una disciplina teórica que formula explicaciones diseñadas para justificar los fenómenos del lenguaje; mientras que en el segundo se refiere al análisis y descripción de los actos lingüísticos; es decir, se encarga del aspecto funcional o uso social del lenguaje. Este segundo aspecto es el que más nos interesa desde el punto de vista del análisis del discurso en la comunicación mediada por el computador. Se trata de un análisis lingüístico que utiliza modelos teóricos del análisis del discurso académico en el marco de la pragmática y la lingüística textual (Cassany, 2006). En el ámbito educativo, este análisis está dirigido a explorar las interacciones, las características y los efectos que generan las actuales prácticas de lectura y escritura en los procesos de comunicación y de enseñanza-aprendizaje en entornos virtuales (Sánchez-Upegui, 2010).

Sánchez-Upegui (2010) realizó un estudio con el propósito de analizar las interacciones de los estudiantes en un foro virtual educativo con el fin de proponer estrategias efectivas de comunicación para estudiantes y docentes.

El foro se realizó en el contexto de un curso de Español, específicamente sobre el tema del plagio académico. La metodología del estudio se inscribe en la lingüística textual cuyo interés se centra en el uso del lenguaje en contexto. Para el análisis de los datos se seleccionaron cuatro perspectivas conceptuales, las cuales fueron: la interacción, la ciberpragmática, la cortesía y la argumentación. La muestra estuvo constituida por 42 estudiantes. Se analizaron 58 mensajes los cuales constituyeron 13.601 unidades léxicas.

Los resultados indican que el foro presentó una dinámica de interacción orientada hacia el estilo monológico de las participaciones, es decir, sin una estructura dialógica o conversacional que llevara al intercambio, con el fin de posibilitar la discusión o el razonamiento grupal. Los mensajes de interacción entre los estudiantes se centraron principalmente en respuestas basadas en la opinión y en actos ilocutivos como la felicitación, el cumplido y el agradecimiento. Esto sugiere que la tendencia de los mensajes fue poco argumentativa con respecto al tema de discusión (el plagio académico). Durante el desarrollo del foro el docente tuvo dos participaciones públicas. Una de ellas fue un mensaje de cortesía orientado a reforzar la imagen positiva de un estudiante, mediante el agradecimiento y el cumplido: “Gracias por su aporte... Aquí se construye y se comparte conocimiento”. La otra participación del docente fue el mensaje inicial, en el cual se informó de manera general el tema del debate: “Los invito a participar con sus opiniones, u otras tomadas de distintas fuentes de información sobre el tema del plagio académico”. El autor formula un conjunto de conclusiones agrupadas en cuatro categorías, a saber: dinamismo comunicativo, corrección lingüística, estrategias regulativas y cortesía; asimismo hace las recomendaciones correspondientes de acuerdo con lo plantado en el propósito del estudio.

### **Enfoque Psicopedagógico**

La investigación educativa que se realiza sobre el proceso de aprendizaje que tiene lugar en los entornos virtuales, en buena medida, es de tipo descriptiva o está orientada hacia el estudio del efecto directo de las TIC sobre el aprendizaje de los estudiantes. Esta estrategia de investigación plantea una relación lineo-causal entre estas tecnologías, asumidas como variable independiente; y su efecto sobre el resultado del proceso de enseñanza-aprendizaje, medido a través de algún indicador de desempeño académico (conocimiento adquirido, satisfacción, cambio de actitud, cambio de conducta) asumido como variable dependiente. Este enfoque de investigación es bastante reduccionista y no hace justicia al nivel de complejidad que dicho proceso comporta.

En consecuencia, abordar el estudio de la interacciones que ocurren en los entornos virtuales de aprendizaje requiere, desde el punto de vista psicopedagógico, al menos, de dos consideraciones previas; en primer lugar, es necesario que el investigador tenga una concepción clara del modelo instruccional implícito, donde estén representados sus tres componentes básicos como son: el estudiante que aprende (E), el docente que diseña y/o gestiona la instrucción (D) y el contenido que debe ser aprendido (C); y al mismo tiempo permita visualizar el rol mediador de las TIC en las interacciones que ocurren entre diferentes pares de estos tres componentes, como sería, por ejemplo, E-TIC-C; E-TIC-D; D-TIC-C. Sin embargo, lo más importante no son las interacciones parciales sino la interacción global de los tres factores mediada por las TIC en el proceso de construcción del aprendizaje (ver Bustos y Coll, 2010).

La segunda consideración tiene que ver con el marco teórico referencial que pudiera ser utilizado para interpretar los procesos que tienen lugar en los entornos virtuales de aprendizaje. En este sentido, existe una marcada

tendencia en la literatura en favor del enfoque sociocultural de la educación representado por Vigotsky (1981). Esta tendencia se fundamenta en el hecho de que el aprendizaje apoyado en las TIC constituye una modalidad instruccional innovadora, creativa y flexible, que privilegia el aprender haciendo, a través del desarrollo de actividades individuales y en grupos, con calidad y pertinencia sociocultural, y la construcción del conocimiento en interacción con otros, bajo la mediación del profesor y/o de los pares.

Esta teoría asume el constructivismo social, o dialéctico, como el soporte epistémico que permite explicar el desarrollo cognitivo humano y el aprendizaje, como parte de un mismo proceso evolutivo. A diferencia de Piaget (1978), quien considera que el desarrollo humano es de naturaleza bio-psicológica o endógeno y que requiere de la disponibilidad en el sujeto de estructuras cognitivas previas para hacer posible el aprendizaje, Vigotsky, por el contrario, considera que lo determinante en el desarrollo humano es la interacción social y la mediación instrumental a través del proceso de aprendizaje. En este sentido, considera que el aprendizaje es el factor principal que impulsa el desarrollo evolutivo-cognitivo del individuo, el cual es el resultado de la mediación social o interpersonal, dando lugar al surgimiento de las funciones cognoscitivas superiores que sirven de soporte a la acción humana, a lo largo de toda la vida.

Desde el punto de vista epistemológico, la teoría sociocultural plantea que el conocimiento no es un objeto que se pasa o transfiere de una persona a otra, sino que se construye por medio de operaciones y habilidades cognoscitivas que se inducen en la interacción social mediada por instrumentos semióticos como el lenguaje y otros sistemas de símbolos y signos. En este sentido, Busto y Coll (2010) señalan que las TIC digitales emergen como potenciales instrumentos psicológicos en el sentido

vygotskiano de la expresión, en tanto que pueden ser usados como mediadores de los procesos intra e intermentales implicados en la enseñanza y el aprendizaje. Estos autores destacan en este planteamiento dos ideas fuerza estrechamente relacionadas, como son: (a) las TIC como instrumentos psicológicos potenciales y (b) las TIC con su capacidad potencial para transformar las prácticas educativas. Tal potencialidad se transformará en acto –en el sentido aristotélico del término- o no, finalmente, en función de los usos efectivos que profesores y alumnos hagan de ellas.

Para explicar como ocurre el proceso de aprendizaje, Vigotsky creó la metáfora de la zona de desarrollo próximo (ZDP), la cual es definida como la distancia que existe entre lo que el aprendiz es capaz de hacer por si solo (sin ayuda), denominado nivel de desempeño real y lo que él podría hacer con ayuda de otro (por ejemplo, el profesor, un compañero con mas experiencia), que sería el nivel desempeño potencial.

Desde el punto de vista educativo la ZDP se convierte en un principio didáctico que guía la instrucción y en un espacio para la intervención pedagógica donde el docente, en su rol de mediador del aprendizaje, diseña la instrucción tomando en cuenta el perfil de entrada del estudiante (conocimientos previos, capacidad, motivación) y ofrece orientación cognitiva, retroinformación y ayuda oportuna para guiarlo hacia el logro de la meta de aprendizaje.

En resumen, la ZDP comporta un proceso construido conjuntamente entre el profesor y el alumno a través del dialogo y la mediación. Este proceso ha sido designado por Wood, Bruner y Ross (1976) como “andamiaje”, entendido como esa estructura exterior desde la que se organiza y se configura una construcción intelectual, política o analítica. El constructivismo sociocultural está presente en la modalidad educativa e-

learning a través de la concepción de un diseño instruccional interactivo y de la función de mediación que ejerce el docente como tutor y gestor del aprendizaje. De acuerdo con esta concepción, se aspira a que el participante aprenda haciendo, compartiendo información y conocimiento, en interacción con personas, contenido y recursos, y en cooperación con sus pares. Este enfoque teórico brinda los fundamentos epistémicos, pedagógicos y psicológicos que permiten diseñar, instrumentar, promocionar, y gestionar la instrucción de acuerdo con los principios del aprendizaje activo, autorregulado y cooperativo, que orientan la acción en los entornos virtuales de aprendizaje.

***El modelo de comunidad de indagación.*** Otro aporte para el estudio de los procesos de aprendizaje mediante las redes electrónicas, desde el enfoque psicopedagógico, lo constituye el modelo de comunidad de indagación (MCI) propuesto por Garrison, Anderson y Archer (2000). Las dos nociones constitutivas juntas, comunidad e indagación, forman un marco organizativo pragmático de los principios y procesos básicos de aprendizaje con el propósito de guiar la práctica de la educación virtual. El modelo se apoya en el desarrollo de una comunidad de aprendizaje, la cual es asumida, por los autores, como un gran reto en el contexto del e-learning.

La fundamentación teórica del MCI se deriva de la concepción de la educación progresiva formulada por Dewey (1933, 1938, 1959), según la cual las nociones de indagación y comunidad constituyen los puntos focales de su filosofía y práctica educativa. Él creyó que toda experiencia educativa debía ser de interés tanto para el individuo como para la sociedad. Para Dewey (1933), la indagación envuelve la generalización del método científico hacia la solución de problemas prácticos que son valiosos para el aprendizaje, lo cual define la relación entre el pensamiento y la acción. Por otra parte, él

destacó la importancia de la comunidad en el desarrollo del individuo a través de la educación y asumió que su esencia estaba en la fusión orgánica del mundo público (social) con el privado (personal). Él creyó que a través de la colaboración los estudiantes asumirían la responsabilidad de construir activamente y confirmar el significado.

De allí que Garrison y Archer (2000) hayan denominado la fundamentación teórica del MCI como constructivismo cooperativo (o transaccional), el cual se basa en el reconocimiento de la estrecha relación que existe entre la construcción personal del significado y la influencia que ejerce la sociedad en la configuración de la relación educativa. Este proceso unificado reconoce la interacción entre el significado individual y el conocimiento socialmente construido. Ellos consideran que la identificación de estos dos intereses es crucial para la construcción de un marco teórico mediante el cual podamos comprender y aplicar el e-learning con fines educativos.

En la propuesta educativa de Dewey se pueden identificar dos principios (a) la interacción, la cual unifica en un marco temporal inmediato los mundos subjetivo (personal) y objetivo (social). Mediante esta interacción se generan las ideas que dan sentido al mundo exterior. Es decir, el significado es construido y compartido. A través de la interacción, las ideas son comunicadas y el conocimiento construido y confirmado; y (b) la continuidad, su importancia radica en que puede constituirse como la base de la educación futura. En este sentido él plantea que “el resultado del proceso educativo debe ser el desarrollo de la capacidad para seguir formándose” (Dewey, 1933, p. 8), de gran valor tanto para el individuo como para la sociedad.

Dewey rechazaba todo pensamiento dualista, sobre todo en relación con

el individuo y la sociedad. Para él el individuo y la sociedad no pueden existir ni por separado ni en una relación de subordinación (Dewey, 1959). Considera que la educación implica comprender esta interacción existente entre la experiencia y los intereses personales, por un lado, y los valores, normas y conocimiento sociales, por el otro. Esta interacción se refleja en la transacción que se produce entre el profesor y el alumno.

La visión transaccional de este enfoque consiste en reconocer que aun cuando el conocimiento sea una construcción social; en un contexto educativo, es el estudiante individualmente el que debe comprender su sentido u ofrecer una comprensión más profunda sobre el mismo. El proceso de conseguir un resultado que sea valioso tanto en el nivel social como personal es la clave de la transacción inherente a la enseñanza. Esta transacción es un elemento común a todas las experiencias educativas incluyendo el e-learning (Garrison y Anderson, 2010).

Por otra parte, Garrison y Cleveland-Innes (2005), consideran que toda experiencia educativa tiene un doble propósito: (a) construir significado (reconstrucción de la experiencia) desde una perspectiva personal; y (b) consiste en afinar y confirmar esa comprensión pero contrastándola en el marco de una comunidad de estudiantes. No obstante, -agregan ellos- si examinamos más de cerca esta transacción, observamos que los roles inherentes tanto a la enseñanza como al aprendizaje son inseparables y percibimos también la importancia de enfocar el proceso educativo como una transacción unificada. Simplemente observamos el mismo proceso desde dos perspectivas diferentes, las cuales plantean cuestiones fundamentales referentes a la responsabilidad y el control del proceso.

El modelo de comunidad de indagación (Garrison, Anderson y Archer, 2000) ha sido desarrollado con el propósito de representar las dimensiones

básicas del aprendizaje en línea como un proceso que emerge de la interacción de tres factores o dimensiones que los autores denominan presencias, a saber: cognitiva, social y docente, las cuales se describen a continuación.

**La presencia cognitiva.** Esta dimensión es definida como “la medida hasta donde el aprendiz es capaz de construir significado a través de la reflexión y un discurso sostenido”. En el MCI, la presencia cognitiva se expresa en cuatro categorías de indagación práctica: (a) evento desencadenante, en la forma de un problema o dilema que requiere solución; (b) la exploración, la cual se enfoca en la búsqueda de información relevante que pueda proveer insight dentro del reto en mano; (c) integración, en la que se establecen conexiones y hay una búsqueda por una explicación viable; y (d) verificación, en la que se selecciona la solución más viable y examina su consistencia. Los atributos de estas categorías son: sentido de desconcierto, intercambio de información y conexión de ideas, y aplicación, respectivamente.

Es importante señalar que estas fases no siguen una secuencia lineal, sino que en cada una de ellas puede haber la necesidad de regresar a etapas previas en búsqueda de nueva orientación o información. La dimensión presencia cognitiva esta basada en el ciclo de indagación reflexiva que debe fundamentar toda experiencia educativa, el cual incluye las fases siguientes: problema, sugerencia, intelectualización, ideas guía, razonamiento resolución y verificación propuesta (Dewey (1933)).

**La presencia social.** Esta dimensión es entendida como la “habilidad de los participantes de una comunidad de indagación para proyectarse a si mismos social y afectivamente, como gente real ... a través del medio de comunicación que está siendo usado” (Rourke, Anderson, Archer, &

Garrison, 1999, p. 10). Tiene el rol de facilitadora y optimizadora del aprendizaje, porque contribuye a la manifestación plena de la presencia cognitiva (Gunawardena, 1995). Es en definitiva la que construye el clima de grupo, aquello que brinda las bases sociales de confianza, calidez, bienestar, empatía y compromiso, necesarias para abrirse a un dialogo comprometido, crítico y sustentable.

En una verdadera comunidad de indagación, el tono de los mensajes es cuestionador pero atractivo, expresivo pero respondiente; escéptico pero respetuoso; desafiante pero solidario (Rourke, Anderson, Archer, & Garrison, 1999). Dentro de esta dimensión se distinguen tres categorías a saber: (a) expresión afectiva; donde el estudiante comparte expresiones de emoción, sentimientos, creencias y valores; (b) comunicación abierta, donde el estudiante construye y mantiene un sentido de compromiso con el grupo; y (c) cohesión del grupo, donde el participante interacciona alrededor de actividades y tareas intelectuales comunes. Los indicadores de estas tres categorías son: auto-proyección/expresión de emociones, confianza/clima libre de riesgo y colaboración/interactividad, respectivamente.

Esta dimensión también está basada en la concepción pedagógica de Dewey (1933), quien pensaba que el aprendizaje era el resultado de la experiencia contextual y socialmente situada. Por lo tanto, la presencia social es vista como un elemento que impacta directamente el desarrollo de la comunidad de aprendizaje y la colaboración de los cursos en línea y, por lo tanto, es una parte integral del modelo MCI.

**La presencia docente.** Esta dimensión es definida como el proceso mediante el cual se hace posible el diseño, facilitación y dirección pedagógica de los procesos cognitivos y sociales con el propósito de obtener resultados de aprendizaje que sean personalmente significativos y educacionalmente

relevantes (Anderson, Rourke, Garrison & Archer, 2001). Este componente del MCI ha sido basado en la concepción educativa de Dewey (1959). El pensaba que el proceso educativo tenía dos lados, a saber: uno psicológico y otro sociológico en los que ninguno de los dos debía estar subordinados el uno al otro. El estableció claramente la necesidad de que el docente tenga un propósito, una estructura y el liderazgo apropiado para conducir el proceso de enseñanza-aprendizaje. Asimismo, argumentó que el profesor tiene la responsabilidad establecer los objetivos y actividades del curso sin necesidad de llegar a ser rígido. Para establecer y mantener una comunidad de indagación, el planteó que el educador debe estar informado, ser flexible pero enfocado y no sentirse incómodo con la incertidumbre. En este sentido, el estableció que “el pensamiento necesita dirección atenta y cuidadosa (1933, p. 22). Dewey, también fue consciente de la necesidad de facilitar relaciones sociales apropiadas prestándole atención a la organización del ambiente social y físico del aula de clase.

Las categorías de esta dimensión son: (a) diseño y organización, el cual no puede ser desatendido en un entorno virtual de aprendizaje. Ello es particularmente cierto en términos de la selección de actividades y tareas relevantes para el aprendizaje colaborativo. En un contexto asincrónico y no-verbal las expectativas con respecto al discurso deben ser claras; por ejemplo, la extensión y foco del mensaje deben ser establecido y modelado; (b) facilitación del discurso; más allá de la consideración sobre que tan clara es la expectativa para la discusión en línea, es necesario guiar la discusión en una manera significativa, garantizando que el estudiante se mantenga enfocado. Es esencial que el estudiante se comprometa y contribuya a la construcción de una comunidad de indagación colaborativa; y (c) instrucción directa; está ampliamente documentada en la literatura especializada la importancia del rol

del docente en el proceso instruccional (ver Anderson et. al., 2001). La presencia docente ha mostrado evidencia de estar conectada al desarrollo del sentido de comunidad en los cursos en línea, como lo evidencian los estudios de Shea, Li y, Swan y Pickett (2005). En un contexto educativo formal, habrá momentos en que se haga necesario intervenir directamente para corregir concepciones erradas, proveer información relevante, resumir la discusión y/o suministrar mediación cognitiva. Los atributos de la categoría presencia docente son: establecimiento del currículo y actividades, modelado de intercambio constructivo y, focalización y resolución de problemas.

## La Observación Flotante en el Ciberespacio

La observación flotante es una técnica de investigación social que data de la década de los años 80. Fue propuesta desde la antropología urbana, para estudiar los procesos dinámicos en la vida de las ciudades, las cuales han sido percibidas por los proponentes de esta técnica, en términos de fluidez, expresada en la movilidad de las relaciones sociales, la fugacidad de las relaciones de tránsito y el anonimato (Hannerz, 1986). Al respecto, Delgado (1999) agrega que se debe “admitir que los objetos potenciales de estudio están en una tupida red de fluidos que se fusionan y licuan, o que se fisionan o se escinden, un espacio de dispersiones e intermitencias, donde los individuos y colectividades se la pasan transitando, circulando, en movilidades, transbordos, correspondencias, con traspiés e interferencias” (p. 45). En este sentido, la observación flotante consiste en estar bajo toda circunstancia vacante y disponible, sin enfocar la atención en un objeto determinado, sino dejarla “flotar” para que no haya filtro, donde no haya a-priori, hasta que aparezcan algunos puntos de referencia y de convergencia, en los que el investigador sea capaz de encontrar ciertas reglas subyacentes (Petonnet, 1982; Neve, 2006).

Por analogía de la dinámica de las ciudades con los flujos de información que ocurren en el ciberespacio, ha habido una exhortación desde las etnografías de lo digital a pensar en metodologías, que tomen en cuenta la fluidez, la multiplicidad, conectividad y segmentatividad de los fenómenos sociales en línea (Escobar, 1994; Dicks y Mason, 1998). Ellos consideran que la observación flotante es una aproximación metodológica que aun cuando fue diseñada para la investigación en las ciudades, debido a la flexibilidad a

la que apela, puede ser de gran utilidad para las etnografías digitales, ya que en el ciberespacio son muchas las opciones que existen para la observación y los aspectos pueden ser exteriorizados a través de las expresiones de un texto, una imagen, un sonido, un video, una figura e incluso hasta mediante los silencios.

Desde la etnografía digital ha surgido, en la última década, una nueva técnica para el estudio de lo social en el ciberespacio, similar a la observación flotante, que recibe el nombre de lucker (del inglés luck, que quiere decir: persona que actúa furtivamente o a escondidas). En este caso, se refiere a aquella persona que visita foros, weblogs, salas de chat, y sitios de Internet sin participar en las discusiones, pasando desapercibido y manteniendo su anonimato (Neve, 2006). Algunos autores han sugerido que el proceso de ser lucker debería ser una fase previa a la observación participante plena, y conllevaría una observación diferida que interprete y negocie los sistemas de significado (Garrido, 2003). No obstante, se ha criticado el hecho de que el lucker, por lo general, adopta una actitud pasiva y que debe ser activo para ser reconocido y realmente poder observar e interpretar las tramas de construcción de sentido (Hine, 2000).

Quienes plantean la observación flotante como una técnica de recolección de datos consideran que se debe distinguir entre la observación flotante que se realiza en entornos digitales y en los no digitales, dado que no existen las mismas condiciones para observar una determinada realidad en un entorno físico en comparación a como se haría en un entorno virtual, siendo esta última más restrictiva. La mayoría de la gente accede al Internet a través de pantallas y no a través de interfaces 3D en 360°; ello hace que el observador tienda a focalizar la mirada sobre una parte de la pantalla, particularmente en aquellos sitios donde el contenido predominante es el

texto escrito. Ellos consideran que este hecho de focalizar la mirada sobre la pantalla tiene el inconveniente que contribuye a que la información sea filtrada por nuestra cultura y sistema de valores (Reguillo, 1998, p. 26).

Frente a la limitación anterior, Neve (2006) propone dos alternativas: (a) leer de manera no direccionada, es decir, no empezar necesariamente por el principio del texto, sino con cierta libertad; por ejemplo, si se está en un foro de discusión en línea se pueden leer fragmentos de algunas contribuciones, subir y bajar la página, ver distintos comentarios con cierta simultaneidad e irse apropiando de ciertos fragmentos de texto; y (b) leer el texto con direccionalidad, atendiendo a posibles polisemias en el discurso, y con sensibilidad, para encontrar en el texto lugares donde se pueda practicar esa observación flotante; de esta forma se podría interpretar el discurso de manera cercana a como el comunicador lo quiso dar a entender.

Otra diferencia entre la observación flotante originalmente propuesta por Pétonnet (1983) y la aplicada al ciberespacio es que en los espacios digitales no se puede observar a las personas y sus acciones como tales, sino mediante sus expresiones discursivas digitalizadas a través de: textos escritos, imágenes, videos, sonidos. Por ello, se recomienda que para observar de manera flotante el ciberespacio se deba:

1. Romper con esquemas metodológicos tradicionales para abordar la observación flotante desde una psicología de lo borroso, otro orden explicativo (...) conceptos borrosos, correlaciones débiles, metáforas que conecten lo describable con lo indescribable una aproximación más plástica que vuelva relevante lo irrelevante (Soto, 1999).

2. Prestar particular atención a lo invisible; es decir, leer entre líneas. “Incluso en espacios no digitales los paisajes son significados de manera

diferente por sujetos distintos”, por lo que surge la pregunta de «para quién es invisible lo que puede ser visible para otros” (Lindon, 2005 p. 3).

3. Poner el énfasis en la observación de aquello que parece banal, cotidiano o aparentemente irrelevante, tomando en cuenta que las interacciones sociales se sustentan en nuevos códigos y lenguajes con los cuales el investigador debe estar familiarizado: los iconos, los emoticonos -símbolos utilizados para expresar estado de ánimo-, los textos, las imágenes y los vídeos, entre otros (Mayans, 2002 y Malinowski, 2007).

4. Atender las letras pequeñas al final de la página, explorar los links fuera de servicio, atender incluso aquellas palabras tan pequeñas o tan grandes que no pueden verse en la pantalla, pues una parte importante de la construcción de significado no es visible... «Lo que se ve oculta lo que no se ve: cuando se contempla algo acabado, no es posible ver las condiciones sociales de su elaboración» (Bourdieu, 1996).

En resumen, en este capítulo se ha hecho una caracterización del escenario de investigación creado por las interacciones humanas que se producen en el ciberespacio y se ha destacado su importancia como fuente de indagación fenomenológica. Asimismo, se presentan las diferentes disciplinas que han contribuido teórica y metodológicamente al estudio del fenómeno educativo que tiene lugar en los entornos virtuales de aprendizaje, como son la antropología, la lingüística y la psicopedagogía. Finalmente, se propone la observación flotante como la técnica de indagación por excelencia para el estudio cualitativo del proceso de enseñanza-aprendizaje basado en las redes de comunicación electrónica, las tecnologías de la información y la comunicación y la Internet.

## *Capítulo 10*

# Análisis de Datos Cualitativos

En los capítulos precedentes presentamos cuatro de las técnicas más utilizadas para la recolección de datos en el proceso de investigación cualitativa en educación; ellas son: la entrevista en profundidad, la observación participante, el grupo focal y la observación flotante. En este nuevo capítulo se describen los procedimientos analíticos básicos que son utilizados en el proceso de reducir los datos y evidenciar su estructura subyacente a objeto de hacerlos interpretables.

El análisis cualitativo se caracteriza por ser interpretativo, en el sentido que facilita la comprensión y explicación de los significados que los actores sociales asignan a las experiencias vividas; es contextual, porque el significado está ligado a un entorno sociocultural específico; es interactivo porque la teoría emerge a medida que se van recolectando los datos, lo cual permite que la misma sea sometida a prueba, sea refinada y re-evaluada.

Diferentes investigadores cualitativos (Taylor y Bogdan, 1996; Rodríguez, Gil y García, 1996; Goetz y LeCompte, 1988; Miles y Huberman, 1984; Strauss y Corbin, 2002; y Martínez, 2004, entre otros); están de acuerdo en que el análisis de datos en este contexto se caracteriza por ser dinámico y recursivo; constituye una parte integral del proceso de investigación. A diferencia del enfoque cuantitativo donde el proceso de análisis constituye una fase específica del proceso secuencial de investigación, posterior a la recogida de los datos, en el enfoque cualitativo el

análisis se realiza a lo largo del proceso de investigación y permite tomar decisiones sobre diferentes cursos de acción en el proceso de investigación.

En el caso de la categorización, el análisis y la interpretación de los contenidos, Martínez (2004) ha señalado que no se trata de actividades mentales separables. Él añade que “nuestra mente salta velozmente de uno a otro proceso tratando de hallarle un sentido a las cosas que examina; se adelanta y vuelve atrás con gran agilidad para ubicar a cada elemento en un contexto y para modificar ese contexto a fondo de acuerdo con el sentido que va encontrando en los elementos” (p. 69). En consecuencia, el análisis cualitativo puede ser definido como un proceso cíclico y circular mediante el cual el investigador organiza y manipula la información recogida para establecer relaciones, interpretar, extraer significados y conclusiones (Spradley, 1980).

## Fases del Análisis Cualitativo

Los datos a ser analizados en la investigación cualitativa pueden estar constituidos por materiales presentados en diferentes formatos, tales como: textos, audio, videos y fotografías, y otros. Pueden provenir de muchos individuos, de pocas personas o de un simple caso. En su procesamiento podemos utilizar diferentes estrategias, por ejemplo: análisis de contenido, análisis del discurso, análisis comparativo constante y análisis estadísticos de datos textuales, entre otros.

En este capítulo, nos centramos en la estrategia del análisis de contenido. En este caso, los datos se caracterizan por ser de naturaleza narrativa o textual que vienen en varias formas y proceden de diferentes fuentes; pueden ser las respuestas de preguntas abiertas a un cuestionario, la transcripción de entrevistas en profundidad o grupo focal, un diario de campo, el texto de un informe ya publicado u otros documentos.

La organización y procesamiento de estos datos no es un trabajo mecánico, sino todo lo contrario, es muy dinámico e interactivo y está cargado de matices interpretativos. De modo que, desde el primer momento en que se empieza a trabajar con los datos, estamos haciendo análisis, avanzamos en el proceso de abstracción y estamos dejando parte de nuestra subjetividad en el proceso de emerger los significados a partir de los datos (González y Cano, 2007).

Los investigadores cualitativos utilizan diferentes procedimientos para procesar sus datos. Así, en algunos casos, como ya hemos dicho, el análisis se realiza en paralelo con el proceso de recolección de datos e información; mientras que en otros los datos se recogen primero y en un momento

posterior se procede a su análisis. Muchos autores prefieren la primera opción porque consideran que este procedimiento hace posible una mayor interactividad entre el investigador, los datos obtenidos y el proceso de análisis. Ello permite ir comprendiendo, desde el primer momento, la situación bajo estudio, validar los datos en el contexto, incorporar detalles recientes relacionados con las percepciones del investigador sobre acciones y hechos, y registrar ideas que vayan emergiendo durante el proceso (memos) para su uso posterior en la interpretación de los datos. De esta forma se puede captar de una manera más rica, variada y sustantiva los detalles de las acciones y los significados que los actores sociales le asignan a aquellas. Esta estrategia es ampliamente avalada por proponentes y desarrolladores del método comparativo constante (ver Glaser y Strauss, 1967; Strauss y Corbin, 2002).

En la segunda estrategia de análisis, los investigadores pueden invertir semanas, meses y aun años en el proceso de recolección de datos para luego iniciar el proceso de análisis. Miles y Huberman (1984) están en desacuerdo con este procedimiento por considerar que el tiempo que media entre la obtención de los datos y el análisis impide recolectar nuevos datos para complementar o profundizar en algún incidente crítico detectado, así como verificar nuevas hipótesis que vayan surgiendo durante el análisis. Igualmente, ellos consideran que esta estrategia tiende a reducir la posibilidad de evaluar hipótesis rivales que cuestionen los sesgos y supuestos asumidos por el investigador, lo que convierte el análisis en una tarea gigantesca que desmotiva y reduce la calidad del trabajo realizado. Por otra parte, existen investigadores cualitativos que para analizar sus datos utilizan los enfoques siguientes:

1. Por preguntas, temas, periodo de tiempo o evento. En este enfoque

se analizan las respuestas dadas por el grupo: (a) por pregunta o tema; (b) durante un periodo de tiempo determinado; o (c) en un evento en particular. En el caso de las preguntas, los datos se organizan tomando en cuenta las respuestas dadas por el grupo a cada una de ellas y se analizan comparativamente a objeto de identificar semejanzas y diferencias. El mismo procedimiento puede ser utilizado con un tema, un periodo de tiempo determinado o un evento en particular.

2. Por caso, individuo o grupo. El análisis se realiza tomando en cuenta cada situación en particular. Un caso pudiera estar referido a una organización, por ejemplo, una escuela, empresa u hospital. Un individuo, como en el caso de un estudiante que participa por primera vez en un programa académico; o por grupo, como ocurriría en una investigación sobre la percepción del trabajo escolar que tienen los docentes recién graduados. También se podría utilizar una estrategia de análisis mixto en el que se combinaran el primer enfoque con el examen de caso, individuo o grupo.

Con un propósito didáctico pudiéramos identificar en el análisis cualitativo las fases siguientes: organización y familiarización con los datos, selección de la unidad de análisis y codificación, categorización/abstracción, presentación de los datos, conclusiones/verificación. Estas fases se describen brevemente a continuación:

***Organización y familiarización con los datos.*** A partir de la recolección de los datos del estudio presentados en sus diferentes formatos que se vayan generando, el investigador debe ir transcribiéndolos y organizándolos para su lectura completa e integradora. Algunos investigadores, después de revisar sus notas de campo, de oír sus grabaciones una y otra vez y de ver las filmaciones realizadas con el propósito de familiarizarse con su contenido y

de revivir las acciones y hechos ocurridos, acostumbran en esta fase ir escribiendo memos que luego ayudarán en el análisis e interpretación de los datos.

En el caso del análisis de contenido con datos textuales, algunos investigadores recomiendan utilizar un diagrama en una página tipo carta donde en el espacio de las tres cuartas parte del lado derecho de la página se pueda vaciar el texto, numerando consecutivamente cada línea y página del mismo. La cuarta parte del espacio restante se usa para colocar las etiquetas que se utilizarán en el proceso de codificación para su posterior categorización y análisis. Para obtener mayores detalles sobre este procedimiento se recomienda revisar a Martínez (2004, p. 73, fig. 5.1.).

***Unidad de análisis y codificación.*** En esta segunda fase, el texto o contenido organizado en la fase anterior se divide en porciones o unidades temáticas, las cuales pueden estar referidas a frases, párrafo o grupo de párrafos que tengan una significación destacable en relación con los objetivos del estudio. En la medida que se van identificando estos elementos de significación (unidades de análisis) se le va asignando un código, nombre o etiqueta que intente compilar el significado emergente. En esta primera etapa, la codificación puede ser de dos tipos, una que algunos investigadores identifican como código “in vivo” cuando el investigador utiliza como etiqueta palabras o frases que aparecen escritas en el texto bajo análisis. El segundo tipo de código es ideado o creado por el investigador; se trata de palabras o frases relevantes que son representativas de los significados que encierran los códigos que agrupa.

Autores como Strauss y Corbin (2002), en el proceso analítico de generar la teoría a partir de los datos, utilizan tres tipos de codificación: abierta, axial y selectiva. En la primera, el investigador identifica las

categorías con sus propiedades y dimensiones. En la segunda, el investigador relaciona las diferentes categorías y subcategorías identificadas previamente en los datos con el propósito de formar explicaciones más precisas y completas sobre el fenómeno bajo estudio. En el tercer tipo de codificación, el investigador intenta integrar y refinar la teoría.

Es recomendable, una vez acabada la asignación de códigos, elaborar un listado de los mismos explicando la significación que el investigador atribuye a cada uno de ellos (a modo de glosario). Esto puede resultar muy útil en el caso de que el investigador se aleje temporalmente de los datos, o si se está compartiendo el trabajo analítico e interpretativo con otros investigadores (triangulación de investigadores). Por otra parte, el esfuerzo de dar una definición a nuestros códigos genera espacios para la reflexión y la gestación de ideas (González y Cano, 2007)

***Categorización/Abstracción.*** La categorización es el proceso mediante el cual se agrupan en clase todos aquellos códigos o etiquetas que comparten un mismo significado. Para ello es importante, a partir de la lectura y relectura del material, identificar temas o patrones (ideas, conceptos, acciones, interacciones, incidentes, terminología o frases) que puedan ser organizados en categorías coherentes. Cada categoría debe ser identificada con una etiqueta que resuma y de significado al texto. La categorización se puede hacer de dos maneras: (a) a partir de una lista de categorías previas derivadas de la teoría de base y/o de la literatura; o (b) fundamentada en los datos. En el primer caso, la categorización consiste en buscar los datos que satisfagan las definiciones de las categorías pre-existentes; mientras que en el segundo, las categorías deben ser construidas con los datos como resultado de develar los patrones, temas o ideas subyacentes en los mismos. Estas categorías deben ser definidas a posteriori del trabajo de agrupar los datos en clase.

En algunos casos estos dos enfoques pueden ser combinados. En este sentido, el investigador puede iniciar el proceso con base en categorías preestablecidas e ir agregando nuevas categorías emergentes. Por otra parte, la lista de categorías iniciales pudiera cambiar en la medida que se avanza en el trabajo de análisis de datos ya que, como hemos dicho antes, se trata de un proceso dinámico e interactivo. Es posible tener que ajustar las definiciones de las categorías o identificar nuevas categorías para acomodar los datos que no encajan en las categorías previas.

Por otra parte, las categorías principales pueden ser descompuestas en subcategorías, lo cual implica reclasificar los datos dentro de cada una de ellas para lograr categorías más pequeñas y precisas, lo que permite una mayor discriminación y diferenciación en los datos. El investigador debe continuar identificando categorías en los datos hasta lograr la saturación de las existentes; es decir, hasta que no aparezcan nuevos temas o subcategorías. Debemos agregar tantas categorías como sean necesarias para reflejar la variación en los datos para una más clara interpretación de los mismos.

En la medida que organizamos los datos en categorías empezamos a observar ciertos patrones y conexiones dentro y entre dichas categorías. Por lo tanto, es importante evaluar la importancia relativa de dichos patrones destacando la variación en sus atributos. Ello se puede hacer de dos maneras:

1. Hacer un resumen descriptivo de los patrones o temas observados. Para ello, es recomendable responder preguntas tales como: ¿Cuáles son las ideas claves expresadas en esta categoría?, ¿Cuáles son las semejanzas y diferencias observadas en los atributos de la categoría?
2. Formar categorías más amplias e inclusivas a partir de las relaciones observadas entre dos o más categorías pre-existentes. Se trata de una categorización de segundo orden o axial donde la nueva etiqueta

integra el significado condensado de los temas o patrones implícitos. Este proceso representa los datos en un mayor nivel de abstracción. A partir de ello se estructura la presentación sintética y conceptualizada de los datos.

***Presentación de los datos.*** Una vez que se ha logrado la organización y comprensión de los datos previamente codificados a través de la búsqueda de relaciones entre diferentes elementos (relación intra grupo) y entre los grupos de significados entre sí (relación entre grupos) procedemos a representar gráficamente dichas relaciones a través de medios tales como: redes de interconexiones, diagramas de flujo, mapas cognitivos u otro. Al respecto, Miles y Huberman (1994), han ensayado procedimientos tales como el de elaboración de matrices, a partir de los cuales, es posible hacer un barrido sistemático de los datos e identificar relaciones, “invisibles” hasta ese momento. En total, estos autores proponen 13 tácticas para generar significación las cuales van desde la simple descripción hasta la explicación, y desde lo concreto hasta lo más conceptual y abstracto.

***Conclusión y verificación.*** Las conclusiones son afirmaciones propositivas en las que se condensa el conocimiento adquirido por el investigador en relación al tema estudiado; refleja una estructura compleja de relaciones que subyace en la información analizada. Las conclusiones del trabajo se empiezan a intuir tempranamente en el proceso del análisis de los datos; en la medida que se avanza en el mismo, el investigador empieza a observar regularidades, patrones y posibles configuraciones que van emergiendo de los datos, pero no es sino al final del análisis cuando las conclusiones pueden ser establecidas con mayor precisión. Este es el momento en que se interpretan los hallazgos obtenidos en la investigación, es decir, se le dice a la audiencia cuál es el significado y la importancia que tienen los mismos.

Para facilitar este proceso es recomendable tomar en cuenta los gráficos y matrices realizados durante la presentación de los datos, a objeto de identificar los conceptos clave y sus relaciones que permitirán describir y explicar los hallazgos emergentes. En este sentido, es importante también reflexionar retrospectivamente acerca de lo que se ha aprendido durante el proceso de investigación, así como sobre las implicaciones teóricas y prácticas de sus resultados.

La verificación, por su parte, consiste en la evaluación crítica de los resultados obtenidos con el propósito de mostrar evidencias de credibilidad de los hallazgos de la investigación. Al respecto, Miles y Huberman (1984) han propuesto trece tácticas que permiten lograr tal propósito, ellas son: chequeo de la representatividad, chequeo de los efectos del investigador, triangulación, ponderación de las evidencias, chequeo de las notas al margen, uso de casos extremos, revisión del significado de los casos o datos extraordinarios (excepciones), revisión de la evidencia negativa, realización de pruebas si... entonces..., resolver y eliminar las relaciones espurreas, replicar hallazgos, chequear explicaciones rivales y obtener la retroalimentación de los informantes (ver Cuadro 39).

*Cuadro 39. Resumen de las Fases del Proceso de Análisis Cualitativo*

Fases	Descripción
<b>Organización y Familiarización con los Datos</b>	Recolección y transcripción de los datos. Organización de la información. Lectura completa e integradora del material.
<b>Unidad de Análisis/ Codificación</b>	Se identifican palabras, frases o párrafos que sean relevantes para los objetivos del estudio.  Se colocan etiquetas o nombre a las unidades de análisis. (Codificación abierta, axial y selectiva.)
<b>Categorización/ Abstracción</b>	Se agrupan las etiquetas o códigos que comparten un mismo significado.  Se asignan nuevas etiquetas que representen categorías emergentes, de segundo grado, a cada uno de los grupos de significados; en esta medida se eleva el nivel de abstracción.
<b>Presentación de los Datos</b>	Descripción de las categorías y subcategorías emergentes. Establecimiento de relaciones lógicas entre ellas. Graficación de tales relaciones.
<b>Conclusiones y Verificación</b>	Son proposiciones en las que se condensa el conocimiento adquirido por el investigador en relación con el tema estudiado; refleja la estructura compleja de relaciones que subyace en la información analizada. Se interpreta el significado e importancia de los hallazgos y se contrasta con el marco teórico referencial.  Para finalizar, es necesario evaluar los hallazgos encontrados. Con este propósito, se utilizan estrategias apropiadas (por ejemplo, triangulación, retroinformación de los participantes) que brinden soporte a la credibilidad del estudio.

## Descripción y Teorización

### Descripción

Los estudios de investigación cualitativa pueden tener un propósito descriptivo o de formulación teórica. El primero tiene como finalidad presentar una caracterización de la realidad estudiada, la cual no es otra que la vida interior de los actores sociales. Al respecto, Emerson (1983) ha señalado que “las descripciones íntimas presentan detalladamente el contexto y los significados de los acontecimientos y escenas importantes para los involucrados” (p. 24). Este tipo de estudio se basa en una síntesis descriptiva de las categorías y relaciones formuladas a partir de la teoría de base que fundamenta la investigación. Un caso típico de análisis descriptivo puede ser observado en los estudios etnográficos, en los cuales el investigador debe hacer un gran esfuerzo por presentar una imagen lo más fidedigna posible de lo que la gente dice y del modo en que actúa; el investigador deja que las palabras y acciones de las personas hablen por sí mismas. Es lo que Martínez (2004) denomina el nivel de descripción normal. Los estudios descriptivos se caracterizan por un mínimo de interpretación y conceptualización. Están redactados de modo tal que permiten a los lectores extraer sus propias conclusiones y generalizaciones a partir de los datos (Taylor y Bogdan, 1996). No obstante, estos mismos autores advierten que:

Sería engañoso sostener que los estudios descriptivos se escriben solos. En todos los estudios los investigadores presentan y ordenan los datos de acuerdo con lo que ellos piensan que es importante. Por ejemplo, en las historias de vida deciden qué incluir y qué excluir, compilan los datos en bruto, añaden fragmentos de conexión entre observadores y

disponen el relato según algún tipo de secuencia. Además, al realizar sus estudios los investigadores toman decisiones sobre lo que deben observar, preguntar y registrar, decisiones que determinan lo que pueden describir y el modo en que lo hacen (pp. 153-154).

Por otra parte, Martínez (op. cit) propone un segundo nivel de descripción que denomina: Endógena. Esta se deriva a partir de las categorías y relaciones que se generan con base en los datos mismos, pero sin llegar a una formulación explícita de la teoría. Al final, la interpretación de los resultados se formula con base en las teorías de base que fundamentan la investigación, lo cual presenta restricciones importantes, como advierte el mismo Martínez cuando expresa: “esto puede tener una seria limitación, especialmente si los datos de los que se extrajeron esas teorías del marco teórico provienen de otros lugares, de otros tiempos, de otras culturas, etcétera” (p. 84).

### **Teorización**

Una teoría es una estructura proposicional que sirve de marco conceptual para comprender, interpretar, explicar y predecir los fenómenos naturales o sociales. Se trata de la expresión máxima del intelecto humano al poder abstraer las regularidades y patrones de comportamientos implícitos en los fenómenos, mediante el uso de un procedimiento sistemático cuyo resultado permite dar cuenta de una parte de la realidad. Esta visión se complementa con la noción de Martínez (2004) quien considera que la teoría “es una construcción mental simbólica, verbal o icónica de naturaleza conjetural o hipotética que nos obliga a pensar de un modo nuevo al completar, integrar, unificar, sistematizar o interpretar un cuerpo de conocimientos que hasta ese momento se consideraban incompletos, imprecisos, inconexos o intuitivos” (p. 88).

En el proceso de la investigación cualitativa la teoría emerge de los datos a través del trabajo sistemático y riguroso en el proceso de análisis que se inicia formalmente con la codificación de los datos y se continúa con la creación y desarrollo de un sistema de categorías interrelacionadas que al ser interpretado permite que la teoría vaya aflorando.

Los estudios de orientación teórica tienen como propósito comprender o explicar rasgos de la vida social que trascienden las personas y escenarios estudiados en particular. En estos estudios los investigadores utilizan los datos descriptivos para ilustrar sus teorías y conceptos y para convencer a la audiencia de la veracidad de lo expuesto. Al respecto, Glaser y Strauss (1967), hacen referencia a dos tipos de teorías: sustantiva y formal. La primera intenta comprender y explicar un caso o situación particular estudiada. Estos autores la refieren como el desarrollo sustantivo o empírico de un área de indagación sociológica, por ejemplo, cuidados de un paciente, relaciones raciales, educación profesional, delincuencia o investigación organizacional.

La teoría formal, por su parte, representa un mayor nivel de abstracción que la teoría sustantiva, lo cual hace que ésta sea más generalizable. Se refiere al desarrollo de una estructura conceptual que intenta explicar un área formal de indagación sociológica, por ejemplo, socialización, autoridad y poder, movilidad social, sistema de recompensa y conducta delictiva, entre otras.

La teorización es el proceso de darle significación a los hallazgos que emergen de los datos durante la interpretación de los resultados de la investigación, lo que al final deviene en una teoría que explica el hecho, fenómeno o situación objeto de estudio. Muchos autores están de acuerdo en que se trata de un proceso complejo, no lineal y recursivo. Martínez (2004)

considera que este proceso “sigue básicamente un movimiento en espiral, del todo a las partes y de las partes al todo, aumentando en cada vuelta el nivel de profundidad, [abstracción] y de comprensión. La visión del todo da sentido a las partes y la comprensión de éstas mejora la del todo: conociendo el bosque se comprenden mejor los árboles, y captando las particularidades de éstos se mejora la comprensión del bosque” (p. 90). Strauss y Corbin (2002) reiteran la idea de la complejidad del proceso de teorización y añaden que el mismo “implica no solo concebir e intuir ideas (conceptos) sino también formularlos en un esquema lógico sistemático y explicativo. Por iluminadora y aun revolucionaria que sea la noción de teorizar, para que una idea llegue a ser teoría se exige que esta se explore a plenitud y se considere desde muchos ángulos o perspectivas” (p. 24).

En resumen se podría decir que la teorización es un proceso de altísimo nivel cognitivo que integra, de una manera compleja, operaciones mentales menores tales como: percibir, comparar, formular preguntas, contrastar, añadir, ordenar, establecer relaciones y especular. Dicho proceso se establece mediante la generación y desarrollo sistemático de las categorías emergentes y sus relaciones en un intento de interpretar el significado de los datos.

## Ejemplo del Proceso de Análisis Cualitativo

En la evaluación del curso Sistema de Mediación Tutorial se administró un cuestionario de autoevaluación con el propósito de conocer la opinión de los participantes (n = 23) sobre diferentes aspectos del curso, cuyo objetivo era desarrollar competencias para llegar a ser un tutor de tesis eficiente; el mismo fue administrado por la Coordinación de Cursos No Conducentes a Título de la UPEL-IPB; fue diseñado y gestionado por el autor de esta obra, bajo la modalidad virtual, en la plataforma Upelvirtual (<http://www.upelvirtual.com/ipb/>), tuvo una duración de 48 horas, y finalizó en abril de 2011. Una de las preguntas del cuestionario era: ¿Que es lo que usted más valora en este curso? Con el propósito de ilustrar el proceso de reducción de datos en el análisis cualitativo hemos decidido procesar la información dada por los participantes a esta pregunta. Para tal fin, procedimos de la manera siguiente:

1. Listamos las respuestas de los 23 estudiantes que respondieron el cuestionario e hicimos una lectura rápida a cada una de las respuestas con el propósito de familiarizarnos con su contenido. Paso seguido, se realizó la enumeración secuencial de las líneas del texto e iniciamos el proceso de segmentación del mismo. Posteriormente, se realizó la codificación usando siglas representativas, como se indica en los Cuadros 40 y 41.

*Cuadro 40. Codificación Abierta*

Código	Estud.	Respuestas a la Pregunta
TUV TIC	1	1. La importancia de un buen docente virtual. 2. El uso de las TIC como medio de conocimiento.
TEC TIC SMT APREL	2	3. La interacción grupal. 4. El uso de la red para comunicarnos. 5. Los procesos cognitivos y metacognitivos para la mediación tutorial. 6. El conocimiento y la aplicación del sistema de mediación tutorial.
SMT APREL	3	7. El estar al tanto que todo proceso tutorial requiere de alcanzar cada una de las submetas del proceso al ritmo del tesista. 8. La adquisición de competencias de investigación que permitan alcanzar la culminación de todo trabajo de investigación.
APREL APREL TIC	4	11. La experiencia de haberlo hecho. 12. Lo que aprendí en el contenido de su programación. 13. La buena plataforma en que el curso fue hecho.
APREL	5	14. Un aprendizaje más que me llena de satisfacción por todos los conocimientos adquiridos. 15. Considero que el estudiante debería hacer primero este curso para luego realizar sus estudio de posgrado. 16. estudio de posgrado.
TUV APREL TEC TEC	6	17. La excelente participación que el profesor tuvo en todo momento con el grupo y en particular con mi persona. 18. Esta experiencia a distancia, para mí, fue exitosa y maravillosa. 19. Formo parte de un equipo muy responsable, siempre dispuesto a trabajar con solidaridad. 20. cooperación y entusiasmo; con humildad y mucho optimismo; aprendimos académicos como equipo. 21. como equipo. 22. como equipo.
SMT SMT	7	23. El dinamismo del proceso, pues cada sección fue graduando su nivel de exigencia. 24. Eso me enseñó a valorar cada actividad y dar lo mejor para estar a la altura del nivel de exigencia. 25. exigencia.
APREL APREL	8	26. El aprendizaje y experiencia que ha dejado en mí como persona, estudiante y tutora iniciada, pues el valor de su contenido es incalculable. 27. iniciada, pues el valor de su contenido es incalculable.
TEC APREL	9	28. Que la interacción virtual para su desarrollo ha sido tan significativa, evidenciado en mi aprendizaje y aplicación práctica de los conocimientos que me ha dejado el deseo de seguirlos realizando. 29. aprendizaje y aplicación práctica de los conocimientos que me ha dejado el deseo de seguirlos realizando. 30. seguirlos realizando.
SMT	10	31. La importancia de la ética que cada tutor de tesis debe poseer en cuanto al ejercicio de sus funciones. 32. sus funciones.
APREL APREL	11	33. El aprendizaje, aunque se me presentaron dificultades. 34. El conocimiento fue lo más valioso que obtuve durante esta experiencia. 35. experiencia.
SMT SMT	12	35. La importancia que el curso da al papel del tutor en el proceso de tutoría de tesis. 36. El saber que un buen tutor no es quien lapida el trabajo del tutorado sino aquel quien ayuda al participante a ser mejor tesista. 37. ayuda al participante a ser mejor tesista.
TUV APREL	13	36. El profesionalismo con que el profesor dio el curso. 38. Las herramientas adquiridas son sumamente importantes para nosotros como tutores. 39. nosotros como tutores.
SMT SMT	14	40. El saber identificar las necesidades y debilidades del tutorado de acuerdo a la matriz COFA, así de acuerdo a ella poder comenzar la mediación tutorial, de una manera como orientador. 41. manera como orientador. 42. manera como orientador.
TUV TEC	15	43. La ayuda y enseñanza del profesor. 44. El compañerismo de los participantes en la interacción de los foros aclarando las dudas o afianzando los planteamientos que allí se hacían. 45. o afianzando los planteamientos que allí se hacían.
APREL TUV	16	46. El aprendizaje obtenido. 47. El haber contactado con la experiencia y conocimiento de un excelente profesor. 48. profesor.
APREL	17	48. Las herramientas obtenidas para mejorar mi desempeño como futuro tutor. 49. futuro tutor.
TEC TUV	18	49. El trabajo grupal de mi equipo. 50. La calidad del profesor, su experiencia y la paciencia que tuvo con los participantes en general y particularmente por los que nunca habían trabajado bajo esta modalidad. 51. general y particularmente por los que nunca habían trabajado bajo esta modalidad.
APREL	19	52. Los aprendizajes que conllevan a mejorar la praxis educativa y tutorial en función de mejorar la calidad educativa que todos estamos comprometidos en llevarla a cabo. 53. mejorar la calidad educativa que todos estamos comprometidos en llevarla a cabo.
SMT	20	54. El permitirme adquirir competencias y capacidades necesarias para llevar a cabo la función tutorial, de una manera reflexiva y crítica, a partir de procesos participativos y dinámicos. 55. función tutorial, de una manera reflexiva y crítica, a partir de procesos participativos y dinámicos. 56. dinámicos.
TEC	21	57. El compartir un espacio donde se pueden plantear dudas, desarrollar debates y procesos de aprendizaje mediante la participación en las actividades propuestas. 58. de aprendizaje mediante la participación en las actividades propuestas.
APREL	22	59. Su contenido. Es muy valioso para aquellos que deseen realizar una tutoría competente. 60. tutoría competente.
TUV	23	60. La objetividad manejada por el profesor, su método de calificación y la claridad en las instrucciones contenidas en las rúbricas de aprendizaje. 61. instrucciones contenidas en las rúbricas de aprendizaje.

*Cuadro 41. Descripción de los Códigos*

<b>Códigos</b>	<b>Descripción</b>
<b>TUV</b>	Tutor Virtual
<b>TIC</b>	Tecnologías de la Información y la Comunicación
<b>TEC</b>	Trabajo en Equipos Colaborativos
<b>SMT</b>	Sistema de Mediación Tutorial
<b>APREL</b>	Aprendizaje Logrado

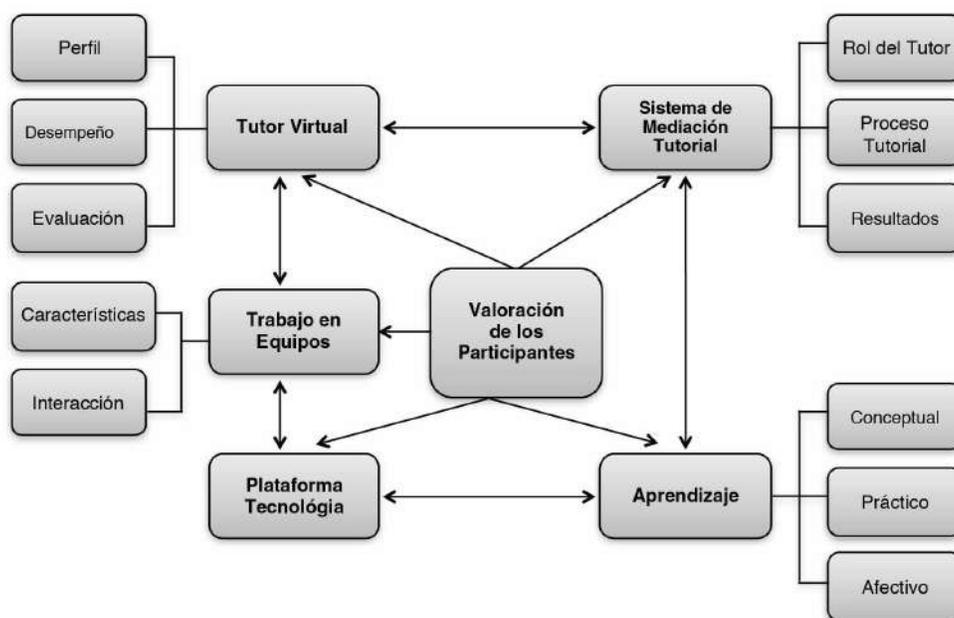
2. En un segundo momento, se agruparon los códigos similares y se formaron las siguientes cinco categorías: Tutor virtual (TUV), Plataforma tecnológica (TIC), Trabajo en equipo colaborativo (TEC), Sistema de mediación tutorial (SMT), Aprendizaje logrado (APREL), como se puede apreciar en el Cuadro 42.

*Cuadro 42. Categorización de acuerdo con similitud de códigos.*

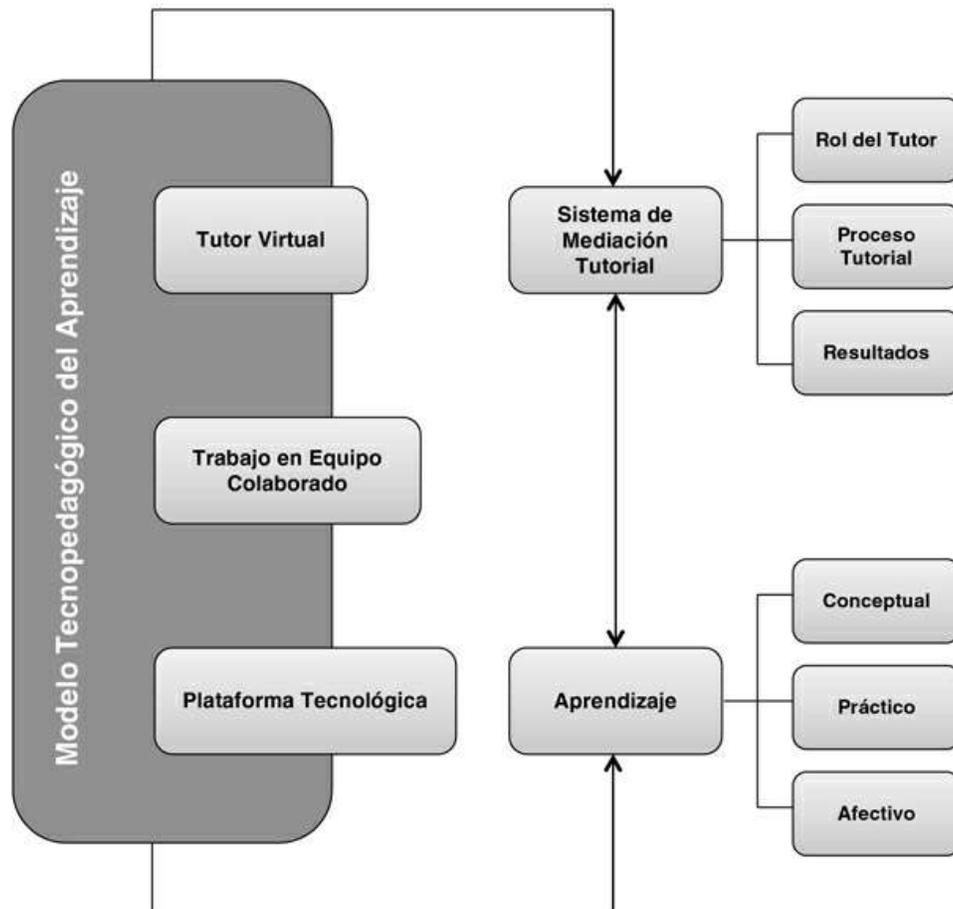
Código	Categoría	Subcategoría	Atributos
<b>TUV</b>	<b>Tutor Virtual</b> Profesional responsable del diseño instruccional virtual y de la gestión del aprendizaje durante el desarrollo del curso.	<b>Evaluación</b>	Método de calificación Rúbricas de aprendizaje Objetividad.
		<b>Desempeño Tutorial</b>	Claridad en las instrucciones. Excelente participación. Ayuda y Enseñanza.
		<b>Perfil</b>	Excelencia, Experiencia, Conocimiento, Paciencia Profesionalismo. Calidad del profesor.
<b>TIC</b>	<b>Plataforma Tecno-educativa</b> Es el entorno virtual que permite la gestión del aprendizaje a distancia.		TIC como medio de conocimiento. TIC como red para comunicarnos. TIC como plataforma instruccional.
<b>TEC</b>	<b>Trabajo en Equipo Colaborativo</b> Equipo de trabajo formado por un número reducido de personas con capacidades complementarias, comprometidas con el logro de un objetivo común, mediante el esfuerzo mutuo y responsabilidad compartida.	<b>Características del equipo</b>	Trabajo en equipo: responsable, dispuesto a trabajar, solidario, cooperativo, con entusiasmo, humildad, optimismo, con acercamiento, compañerismo.
		<b>Interacción</b>	Interacción virtual, interacción grupal, interacción en los foros, compartir, espacios, plantear dudas, debates, procesos de aprendizaje.
<b>SMT</b>	<b>Sistema de Mediación Tutorial</b> Representa el contenido y objetivo principal del curso, el cual estaba referido al desarrollo de competencias para ejercer la función de tutor de tesis de grado de manera eficiente.	<b>Rol del tutor</b>	La importancia rol del tutor Buen tutor ayuda a ser mejor tesista. Tutor como orientador. Función tutorial reflexiva y crítica Mediación tutorial Uso de matriz DOFA.
		<b>Proceso tutorial</b>	Uso de procesos cognitivos y metacognitivos Las submetas del proceso al ritmo del tesista El dinamismo del proceso graduado al nivel de exigencia.
		<b>Resultados</b>	Logro de competencias y capacidades para la función tutorial. Competencia para la investigación.
<b>APREL</b>	<b>Aprendizaje</b> Se refiere a los cambios internos y externos del estudiante en cuanto a su nivel de competencia para desempeñarse como un tutor de tesis de grado.	<b>Conceptual</b>	Contenido programático del curso Psicología cognitiva. Ética del tutor.
		<b>Practico</b>	Adquisición de competencias de investigación. Adquisición de herramientas. Experiencia como persona, estudiante y tutor. Aplicación práctica del conocimiento.
		<b>Psicoactivo</b>	Deseo de seguirlos realizando las actividades del curso Aprendizaje que me llena de satisfacción. Experiencia exitosa y maravillosa.

3. Una vez realizada la primera fase de la categorización, se procedió a realizar un análisis dentro de cada categoría con el propósito de identificar sub-categorías, como se puede apreciar en el cuadro 42, antes mencionado.

4. Se presentaron los resultados del análisis mediante dos gráficos que representan las categorías y subcategorías identificadas y sus relaciones. En el grafico 6 se integran las categorías: Tutor virtual, Plataforma tecnológica y Trabajo en equipo colaborativo en una categoría superior que hemos denominado modelo tecno-pedagógico. Esta nueva categoría se vincula con el sistema de mediación tutorial y el aprendizaje logrado en el curso (ver Figuras 5 y 6).



*Figura 5. Resultados del análisis de datos agrupados en categorías y subcategorías*



*Figura 6. El modelo tecnopedagógico y sus relaciones*

5. Finalmente, se hizo la descripción e interpretación de los resultados tomando en cuenta el planteamiento de la pregunta cuyas respuestas fueron analizadas en el presente ejemplo.

### **Descripción de los Resultados**

Como ya se indicó, el análisis de los datos se basó en la información contenida en las respuestas a la pregunta ¿Qué es lo que usted más valora en este curso?, la cual formó parte del cuestionario de evaluación que se

administró a los participantes al final del curso Sistema de Mediación Tutorial. De acuerdo con el DRAE (Diccionario de la Real Academia Española), valorar significa reconocer, estimar o apreciar el grado de utilidad que tiene una persona para satisfacer las necesidades o proporcionar bienestar o deleite. En este caso nos interesaba indagar en qué medida los participantes eran capaces de reconocer la utilidad del curso para su formación como futuros tutores de tesis de grado. El análisis permitió develar la existencia de cinco categorías subyacentes en los datos, ellas fueron: tutor virtual, plataforma tecnológica, trabajo en equipo, sistema de mediación tutorial y aprendizaje logrado.

**Tutor virtual.** Esta categoría se refiere al profesional responsable del diseño instruccional virtual y de la gestión del aprendizaje durante el desarrollo del curso. Esta denominación es análoga a la del instructor de un curso académico presencial. El análisis de los atributos de la categoría permitió identificar tres subcategorías, a saber:

1. El perfil del tutor, referido a las características profesionales y personales del profesor del curso; dentro de estas fueron identificadas las de: excelencia, experiencia, conocimiento, profesionalismo, paciencia y calidad del profesor. Un ejemplo de como se reflejo esta subcategoría en las respuestas de los estudiantes fue cuando el informante No. 18 señaló [lo que más valoré fue]: “la calidad del profesor, su experiencia y la paciencia que tuvo con los participantes en general y particularmente por los que nunca habían trabajado bajo esta modalidad”.

2. El desempeño tutorial, estaba referida a las actividades y acciones del profesor del curso en su rol de gestor, orientador, animador y apoyo a los participantes durante el desarrollo del curso. Al respecto, se identificaron los atributos siguientes: claridad en las instrucciones,

excelente participación, ayuda oportuna, calidad de la enseñanza. Un ejemplo de como se reflejo esta subcategoría en las respuestas de los estudiantes fue cuando el informante No. 6 expresó: “lo que más valore en este curso, es la excelente participación que el tutor tuvo en todo momento con el grupo y en particular con mi persona; esta experiencia a distancia, para mi, fue exitosa y maravillosa”.

3. La evaluación, referida a los procedimientos utilizados por el profesor del curso para valorar el aprendizaje logrado por los participantes durante y al final del curso. Los atributos identificados fueron: la objetividad en el proceso de evaluación, el método de calificación utilizado y el uso de la rúbrica como instrumento de evaluación. Un ejemplo de como se reflejo esta subcategoría en las respuestas de los estudiantes fue cuando el informante No. 23 señaló [lo que más valoré fue]: “la objetividad manejada por el tutor, su método de calificación y la claridad en las instrucciones contenidas en las Rúbricas [de evaluación del] Aprendizaje”.

**Plataforma tecnológica.** Constituye el entorno virtual desde donde se gestiona el aprendizaje. En el caso del curso Sistema de Mediación Tutorial, el curso fue diseñado en la plataforma Moodle de la UPEL-IPB, conocida como upervirtual, siguiendo los procedimientos de la metodología PACIE (Camacho, 2008), la cual permite una orientación apropiada para que el participante pueda navegar a través de la plataforma para el desarrollo de las diferentes actividades de aprendizaje. Además, hace posible la presentación didáctica del contenido en diferentes formatos (texto, video, blog, ppt, website) desarrollo de tareas (interacción individual con el contenido, discusión en foros, trabajo en equipos colaborativos a través de la wiki, taller

de dudas, entre otras), evaluación en línea mediante cuestionarios, rubrica, cumplimiento de tareas, portafolio y otros.

La plataforma también permitió la información permanente del participante a través de una cartelera informativa y del sistema de mensajería interna. Asimismo, se promovió la interacción social informal mediante el uso de una cafetería virtual o CyberCafe. Los atributos identificados en esta categoría fueron: las TIC como plataforma instruccional, como medio de conocimiento y como red de comunicación. Algunos ejemplos de como se reflejo esta categoría en las respuestas de los estudiantes fue cuando los informantes señalaron [lo que más valoré fue]: “el uso de las TIC como medio de conocimiento” (Informante No. 1); “el uso de la red para comunicarnos” (Informante No. 2) y “la buena plataforma en la que el curso fue hecho” (Informante No. 4).

***Trabajo en equipo colaborativo.*** Un equipo de trabajo puede ser definido como un número reducido de personas con capacidades complementarias, comprometidas con el logro de un objetivo común, mediante el esfuerzo mutuo y responsabilidad compartida. En el caso el curso SMT, el 50 % de las tareas de aprendizaje fueron realizadas en equipos colaborativos de cuatro o cinco participantes. Ello implicaba una organización interna en cada equipo que empezaba por la designación de un coordinador, establecer un cronograma de trabajo, asignar responsabilidades, definir mecanismos y frecuencia de comunicación, la entrega de un informe final y/o la discusión de dicho informe con el profesor del curso mediante el programa Skype.

El análisis de los atributos de la categoría permitió identificar dos subcategorías a saber: características del equipo e interacción en grupo. Los atributos más relevantes que definen la primera subcategoría son: trabajo en

equipo, responsable, dispuesto a trabajar, solidario, cooperativo, con entusiasmo, humildad, optimismo, con acercamiento y compañerismo. En cuanto a la subcategoría de interacción los atributos fueron: interacción virtual, interacción grupal, interacción en los foros, compartir espacios, plantear dudas, debates, procesos de aprendizaje. . Algunos ejemplos de como se reflejo esta categoría en las respuestas de los estudiantes fue cuando los informantes señalaron [lo que más valoré fue]: “la interacción grupal” (Informante No. 2). Por otra parte, el informante No. 6 expreso: “forme parte de un equipo muy responsable, siempre dispuesto a trabajar con solidaridad, cooperación y entusiasmo... con humildad y mucho optimismo... aprendimos a acercarnos como equipo”.

***Sistema de Mediación Tutorial.*** Esta categoría representa el contenido y objetivo principal del curso, el cual estaba referido al desarrollo de competencias para ejercer la función de tutor de tesis de grado de manera eficiente. El SMT constituye un modelo de tutoría, desarrollado por Ruiz Bolívar (2006), centrado en la mediación cognitiva del tutor como estrategia para lograr que el estudiante desarrolle una buena tesis y adquiere las competencias de un investigador.

El análisis de los atributos de esta categoría permitió develar las siguientes sub-categorías, las cuales fueron: (a) rol del tutor, referido a las funciones que desempeña un tutor de tesis de grado; los atributos que definieron esta subcategoría fueron: función tutorial reflexiva y critica, mediación tutorial, tutor como orientador, el buen tutor ayuda a ser mejor tesista y la importancia del tutor; (b) proceso tutorial comprende las diferentes acciones, interacciones y transacciones que comparten el tutor de tesis con sus estudiantes durante el desarrollo del trabajo de grado. Este

proceso comprende tres momentos, a saber: la situación inicial, el proceso de transformación y la situación final.

Los atributos que destacaron en esta subcategoría fueron; el uso de la matriz DOFA como herramienta de diagnóstico del perfil académico del estudiante en el momento de la situación inicial, la mediación de procesos cognitivos y metacognitivos, el logro de cada uno de los 12 subtemas que comprende el modelo, tomando en cuenta el ritmo de aprendizaje del tesista, el dinamismo del proceso graduado al nivel de exigencia; y (c) resultados, referidos a los logros que se espera debe tener el estudiante tesista una vez que finaliza su tesis siguiendo el modelo SMT. Los atributos identificados fueron: logro de competencias y capacidades para la función tutorial y competencia para la investigación.

Un ejemplo de como se reflejó esta categoría en las respuestas de los estudiantes fue cuando el informante No. 7 señaló [lo que más valoré fue]: “la importancia que el curso da al papel del tutor en el proceso de tutoría de tesis...El saber que un buen tutor no es quien lapida el trabajo del tutorado sino aquel quien ayuda al participante a ser mejor tesista”.

**Aprendizaje.** Esta categoría representa los logros obtenidos por los participantes al final del curso; es decir, se refiere a los cambios internos y externos del estudiante en cuanto a su nivel de competencia para desempeñarse como un tutor de tesis de grado. El análisis de los atributos de categoría permitió develar la existencia de tres sub-categorías, a saber: (a) conceptual, se refiere al aprendizaje de tipo teórico, informativo obtenido por el estudiante durante el desarrollo del curso. Los atributos más destacados fueron: contenido programático del curso, psicología cognitiva y ética del tutor; (b) practico, incluye el tipo de aprendizaje procedimental o saber hacer que logran los estudiantes durante el curso. Los principales atributos fueron:

adquisición de competencias de investigación, adquisición de herramientas, experiencia como persona, estudiante y como tutor, aplicación práctica del conocimiento; y (c) psico-afectivo, la cual toma en cuenta los aprendizajes no-cognitivos adquiridos por los participantes durante el curso. Sus atributos fueron: logré un aprendizaje que me llena de satisfacción, fue una experiencia exitosa y maravillosa, deseo seguir realizando las actividades del curso. Algunos ejemplos de como se reflejo esta categoría en las respuestas de los estudiantes fue cuando los informantes señalaron [lo que más valoré fue]: “un aprendizaje más que me llena de satisfacción por todos los conocimientos adquiridos” (Informante 5); “el aprendizaje y experiencia que ha dejado en mi como persona, estudiante y tutora iniciada, pues el valor de su contenido es incalculable” (Informante 8); “las herramientas obtenidas para mejorar mi desempeño como futuro tutor” (Informante 17).

### **Conclusión y Verificación**

Al presentar gráficamente los resultados del análisis nos damos cuenta no sólo de las categorías y subcategorías emergentes, sino también de las relaciones existentes entre ellas. Es así como, a partir de las categorías tutor virtual, plataforma tecnológica y trabajo en equipo colaborativo, hemos podido generar una nueva categoría denominada modelo tecno-pedagógico de aprendizaje el cual hipotetizamos está directamente relacionado con las categorías sistema de mediación tutorial y aprendizaje logrado, en los términos siguientes: cuando los participantes respondieron a la pregunta ¿Qué es lo que más valoras de este curso?, implícitamente se estaban refiriendo al merito del modelo instruccional utilizado, que hemos denominado tecno-pedagógico, al hacer posible aprehender los contenidos del curso (SMT) a objeto de lograr los resultados de aprendizaje señalados (conceptual, practico y psico-afectivo).

En conclusión, lo que los participantes valoraron más del curso Sistema de Mediación Tutorial, fue: (a) el profesor del curso, el cual actuó como tutor virtual del mismo y de éste valoraron su perfil profesional, el desempeño como docente del curso y el sistema de evaluación empleado; (b) la plataforma tecnológica, como entorno de aprendizaje virtual; (c) el trabajo en equipo colaborativo, como un recurso eficiente para la construcción social del aprendizaje, especialmente en lo referente a sus características y la interacción entre sus miembros; (d) el sistema de mediación tutorial, como objeto central del curso; en tal sentido, valoraron el rol que en el modelo SMT se asigna al tutor de tesis, el proceso tutorial como tal y los resultados del mismo; y (e) el aprendizaje logrado, como resultado de la experiencia del curso; valoraron particularmente el aprendizaje conceptual, práctico y psicoactivo.

Si tuviéramos indagando sobre un problema de investigación real, en lugar de estar respondiendo hermenéuticamente una pregunta evaluativa sobre un curso, como ocurre en esta ocasión, correspondería en esta sección hacer la discusión de los hallazgos encontrados en los datos a la luz de la literatura revisada o el marco teórico referencial de la investigación. Por ejemplo, en un tema como este pudiera ser plausible vincular la interpretación de los datos a la teoría sociocultural del aprendizaje propuesta por Vigotsky (1981).

Esta circunstancia de que se trata básicamente de un ejercicio de análisis de datos más que de una investigación real, limita las posibilidades de que se pudiera cumplir también con la fase de verificación contemplada en el esquema para el análisis de datos cualitativos. No obstante, el ejercicio cumple una función didáctica al guiar al investigador novel por la secuencia

lógica de las actividades principales del proceso de reducción de datos en este tipo de enfoque investigativo.

## Analisis Cualitativo Computarizado

El auge de la investigación cualitativa en educación evidenciado desde la década pasada, por una parte, y el avance de la tecnología de la información y la comunicación logrado durante los últimos años, por la otra, han hecho posible que programadores informáticos hayan desarrollado diferentes software para el análisis de datos cualitativos presentados en formato de texto, gráfico, audio o video. Estos programas tienen entre otras ventajas el que permiten al investigador manejar un mayor volumen de datos, en comparación con el procesamiento manual, con mayor precisión, seguridad y rapidez. Sin embargo, es muy importante aclarar que el uso de este tipo de programas, por sí mismo, no es garantía de calidad en la investigación. Un estudio deficientemente diseñado, con datos mal recogidos, parciales o sesgados no mejorara sus resultados por el simple hecho de utilizar una herramienta tecnológica para el análisis de los datos, por muy eficiente que ésta sea.

El uso de este tipo de programa de análisis nos permite (a) codificar segmentos de texto en una base de datos para posteriormente consultar todos los segmentos identificados con el mismo código; (b) recuperar texto por patrones o estructuras de hipótesis que incluyan la relación de casos positivos y negativos de un código; (c) realizar búsquedas textuales con autocodificación de los resultados; (d) guardar los resultados de las búsquedas textuales y la posibilidad de que las salidas de los análisis parciales de las operaciones aplicadas sobre los datos pasen a formar parte del mismo sistema, convirtiéndose en datos primarios para ser utilizados en análisis posteriores; (e) trabajar con una sola copia de cada documento,

guardando las direcciones de los segmentos organizados en una jerarquía arborescente; y (f) diseñar diagramas, redes semánticas o mapas conceptuales en los que se pueden evidenciar las relaciones implícitas en los códigos o categorías, para su posterior interpretación (Rodríguez,1997). Algunos de los programas más conocidos son: ATLAS.ti, N-Vivo, NUD\*IST, WINMAX, ETNOGRAPH 6 y MAXQDA 11 (ver reseña en el Cuadro 42).

Estos programas funcionan bajo el régimen de software propietario; es decir, que se debe pagar para tener acceso a los mismos. Su adquisición por parte de estudiantes está sujeta a una tarifa reducida. No obstante, la mayoría de ellos mantienen dentro de su política de uso, el acceso temporal a los mismos (15 a 30 días) en forma gratuita, con un propósito de prueba y evaluación. Asimismo, para facilitar la familiarización de los investigadores con los programas éstos disponen de tutoriales y sistemas de ayuda para la capacitación y entrenamiento de quienes acceden a ellos por primera vez. En el Cuadro 43 se han dejado las correspondientes direcciones electrónicas para facilitar el acceso a los mismos por parte del lector interesado.

*Cuadro 43. Programas Electrónicos para el Análisis de Datos Cualitativos*

Programa	Características	URL
<b>ATLAS.Ti</b>	Procesa los formatos más comunes de datos textuales, gráficos, audio y video. Maneja más de 20 formato gráficos, formatos de audio como: WAV, MP3, WMA; así como formatos de video, por ejemplo: AVI, MPG y WMV. Permite: (a) usar redes para explorar y descubrir las relaciones subyacentes en la información obtenida; (b) construir teoría basada en los datos; (c) recuperar información y navegar por todos los segmentos de datos.	<a href="http://www.atlasti.com">www.atlasti.com</a>
<b>N-Vivo 9</b>	Es de gran ayuda en la organización, clasificación y análisis rápido de la información. Permite: (a) explorar y visualizar patrones y compartir su trabajo con otros; (b) descubrir relaciones sutiles en los datos; (c) intercambiar datos con aplicaciones como: Microsoft Excel, Word, IBM-SPSS, Statistics y EndNote.	<a href="http://www.qsrinternational.com/other-languages_spanish.aspx">http://www.qsrinternational.com/other-languages_spanish.aspx</a>
<b>NUD*IST</b>	Permite dividir la información textual, asignar categorías las cuales pueden ser agrupadas y organizadas jerárquicamente, establecer relaciones entre categorías y realizar búsquedas textuales específicas, construir matrices y tablas de frecuencia. Hace posible identificar las tendencias subyacentes en los datos y extraer conclusiones relevantes.	<a href="http://www.qsrinternational.com/products_previous-products_nvivo9.aspx">http://www.qsrinternational.com/products_previous-products_nvivo9.aspx</a>
<b>ETNOGRAPH 6</b>	Hace posible: (a) crear proyecto de análisis y añadir datos; (b) codificar, editar y recodificar archivos; (c) añadir y codificar un segundo archivo de datos; (d) buscar información por segmento codificado; (e) explorar satos y lista de memos; (f) explorar relaciones ente códigos y hacer tablas de frecuencia.	<a href="http://www.qualisresearch.com/default.htm">http://www.qualisresearch.com/default.htm</a>
<b>MAXQDA 11</b>	Permite organizar, evaluar e interpretar sus datos. Crea fácilmente soporte visual y permite conectar y compartir con otros investigadores. Posibilita la codificación gráfica mediante el uso de más de 300 emoticones. Puede codificar audio y video. Elabora tablas de frecuencia y las grafica. Puede importar tablas como documento: importa directamente archivos XLS, XLSX, los codifica, edita y los analiza como si fueran archivos de texto.	<a href="http://www.maxqda.com/products/maxqda11">http://www.maxqda.com/products/maxqda11</a>

## Referencias

- Aaker, D. A., y Day, G. S. (1989). *Investigación de Mercado*. México: MacGraw-Hill.
- Adams, G. S. (1970). *Medición y Evaluación*. Barcelona (España): Herder.
- Adorno, T. W., Frenkel, E., Levinson, D. J., y Sanford, R. N. (1969). *The authoritarian personality; Studies in prejudice*. New York: Norton.
- Albornoz, O. (1987). *La educación en el estado democrático*. (2da. Ed.). Ediciones de la Biblioteca UCV.
- Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment*, 2, 1, 15-25.
- Allport, G. W. (1935). *Attitudes*. En C. Murchinson (ed). A handbook of social psychology. Worcester. Mass: Clack University Press.
- Anastasi, A. (1976). *Psicological testing*. New York: McMillan Publishing, Co.
- Ander-Egg, E. (1995). *Técnicas de investigación social*. Buenos Aires: LUMEN 24.
- Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks*, 5(2), 1-17.
- Armony, V. (1997). El análisis de datos cualitativos en ciencias sociales: Nuevos enfoques y herramientas *Revista de Investigaciones Folklóricas*, N° 12, p. 9-16.
- Ausubel, D. Novak, J. D., y Hanesian, H. (1983). *Psicología educativa: Un enfoque cognitivo*. Mexico: Trillas.
- Baker, J. A. (1995). *Paradigmas: El negocio de descubrir el futuro*. Bogota: McGraw-Hill Interamericana.
- Banco Mundial. (1993). *Venezuela en el año 2000*. Educación para el crecimiento económico y

equidad social. Vol. I. Caracas: Autor. Mimeo.

Berkowitz, L. (1972). *Social psychology*. Blenviews, Ill.: Scott Foreman.

Berdie, D. R., y Anderson, J. F. (1974). *Questionnaires: Design and use*. Metuchen, N. J.: The Scarecrow, Press Inc.

Bericat, E. (1998). *La integracion de los metodos cuantitativos y cualitativos en la investigacion social*. Barcelona: Ariel.

Bingham, W. V. D. y Moore, B. V. (1973). *¿Cómo entrevistar?*. Madrid: Rialp.

Black, J. A., y Champion, D. J. (1976). *Methods and issues in social research*. New York: Wiley and Sons Inc.

Blalock, H. M. (1977). *Estadística Social*. México: Fondo de Cultura Económica.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Bloom, B. S., Hastings, T., y Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Blumer, H. (1982). *El interaccionismo simbolico: perspectiva y método*. Barcelona: Hora.

Bollen, K., y Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin* N° 110, 305-314.

Bonoma, T. (1985) "Case Research in Marketing: Opportunities, Problems and a Process", *Journal of Marketing Research*, 2(2), pp. 199-208.

Bourdieu, P. (1996). *Sobre la televisión*. Barcelona: Anagrama.

Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71, (6), 733-750.

Bunge, M. (2000). *La investigación científica*. Barcelona: Ariel.

- Brown, F. G. (1980). *Principios de la medición en psicología y educación*. México: El Manual Moderno.
- Burt, C. L. (1966). Appropriate uses of factor analysis and analysis of variance. En R. B. Cattell (Ed.) *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.
- Bustos, S. A., y Coll, S. C. (2010). *Los entornos virtuales como espacios de enseñanza y aprendizaje. Una perspectiva psicoeducativa para su caracterización y análisis*. Disponible en: [http://www.scielo.org.mx/scielo.php?pid=S1405-66662010000100009&script=sci\\_arttext](http://www.scielo.org.mx/scielo.php?pid=S1405-66662010000100009&script=sci_arttext). Consulta: Octubre 12, 2012.
- Buxarrais, M. R. (1989). Análisis de la interacción profesor-alumno como catalizadora del proceso de aprendizaje. *Revista de Educación*, No. 288, pp. 419-428. Universidad de Barcelona.
- Camacho, P. (2008). *Metodología PACIE*. Disponible: [http://vgcorp.net/pedro/?page\\_id=20](http://vgcorp.net/pedro/?page_id=20). Consulta realizada: Septiembre 27, 2012.
- Campbell, D. T. (1963). Experimental and quasi-experimental designs for research on teaching. En N. L. Gage (ed). (1963). *Handbook of research on teaching*. Chicago: Rand McNally.
- Campbell, D. T., y Fiske, D. W. (1959). Convergent and discrimination validation by the multitrait multimethod matrix. *Psychological Bulletin*, 56,81-105.
- Cardno, J. A. (1955). The notion of attitude: An historical note. *Psychology Rep.* 1,345-352.
- Cronbach, L. J. (1963). Course improvement through education. *Teacher College Record*. 64, 672-683.
- Cronbach, L. J. (1960). *Essentials of psychological testing*. New York: Harper and Row.
- Cassany, D. (2006). *Tras las líneas*. Barcelona: Anagrama.
- Catterall, M., y Maclaran, P. (1997). *Focus group data and qualitative analysis*. Sociological Research Online. Disponible: <http://www.socresoline.org.uk/2/1/6.html>. Consulta: Agosto 7, 2012.

- Cattell, R. B. (1952). *Factor Analysis*. New York: Harper.
- Champion, D., y Sear, A. (1969). Questionnaire response rate: A methodologic analysis. *Social Forces*, 47, 335-339.
- Chase, C. I. (1968). The impact of some obvious variables on essay test scores. *Journal of Educational Measurement*, 5, 315- 318.
- CINTERPLAN. (1987). *Innovación educativa y formación de docentes: Implicaciones en el planeamiento de la educación*. Caracas: Autor.
- Coles, R. (1971). *Migrants sharecroppers mountaineers*. Boston: Little Brown.
- Coll, C., y Sole, I. (1990). La interacción profesor-alumno en el proceso de enseñanza- aprendizaje. En C.
- Coll, J. Palacios, y A. Marchesi (eds). *Desarrollo psicológico y educación*. Vol II. Psicología de la Educación. Madrid: Alianza Editeorial, pp. 315-325.
- Comisión Presidencial para la Reforma del Estado. (1990). *Un proyecto educativo para la modernización y la democratización*. Caracas: Autor.
- Conradson, D. (2005). Focus groups. In R. Flowerdew y D. Martin (Eds.), *Methods in human geography*. (pp. 128-143). Harlow, UK: Pearson.
- Consejo Nacional de Educación. (1993). *Plan decenal de educación*. Edición especial. Versión preliminar. Caracas: Autor. Mimeo.
- Cook, S. W., y Selltiz, C. (1964). A multi indicator approach to attitude measurement. *Psychological Bulletin*, 62, 36-55.
- Cook, T. D., y Reichardt Ch. S. (1986). *Metodos cualitativos y cuantitativos en investigacion evaluativa*. Madrid: Morata.
- Coopersmith, S. (1959). A method for determining self-esteem. *Journal of abnormal and social*

*psychology*, 59,87-94.

Cronbach, L. J. (1963). Course improvement through education. *Teacher College Record*, 64,672-683.

Cronbach, L. J. (1960). *Essentials or psychological testing*. New York: Harper and Row.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of the test. *Psychometrika*, 16, 297-334.

Danhke, G. L. (1989). Investigación y comunicación. En C. Fernandez-Collado y G. L. Danhke (comps). *La comunicación humana: ciencia social*. Mexico: McGraw-Hill, 385-454.

Dawes, R. M. (1975). Graduate admissions variables and future success. *Sciences*, 187,721-723.

Dawson, J. B., y Thomas, G. H. (1972). *Item analysis and examination statistics*. Birmingham: Union of Educational Institutions.

Delgado, M. (1999). *El animal público, Hacia una antropología de los espacios Públicos*. Barcelona: Editorial Anagrama.

Dewey, J. (1933). *How we think* (rev ed.). Boston: D. C. Heath.

Dewey, J. (1938). *Experience and education* (7th printing, 1967). New York: Collier.

Dewey, J. (1959). My pedagogic creed. In J. Dewey, *Dewey on education* (pp. 19-32). New York: Teachers College, Columbia University. (Original work published 1897).

Díaz-Barriga, A. (1987). Problemas y retos del campo de la evaluación educativa. *Perfiles Educativos*, 37, 3-15.

Dicks, B., y Mason, B. (1998). Hypermedia and Ethnography: Reflections on the Construction of a Research Approach. *Sociological Research Online*, vol. 3, no. 3. Disponible: <http://www.socresonline.org.uk/socresonline/3/3/3.html>. Consulta: Octubre 22, 2012.

- Dilthey, W. (1972). *The rise of hermeneutics: 1900*. New Literary History.
- Douglas, J. D. (1976). *Investigative social research: Individual and field team research*. Beverly Hills (California): Sage Publications.
- Doyle, W. (1992). *Research on classroom organization and management*. Invited paper prepared for US Department of Education. Washington, D. C.
- Downey, G.L., Dumit, J. & Williams, S. (1995). Cyborg Anthropology. *Journal of Cultural Anthropology*, 10 (2), pp. 265-269.
- Ebel, R. L. (1977). *Fundamentos de la medición educacional*. Buenos Aires: Guadalupe.
- Edwards, A. L. (1957). *Techniques of attitude scale construction*. New Jersey: Prentice-Hall, Inc.
- Edwards, A. L., y Kilpatrick, F. P. (1948). A technique for the construction of attitude scale. *Journal of Applied Psychology*, 32, 374-384.
- Emerson, R. M. (1983). (Comp.). *Contemporary field research*. Boston: Little Brown.
- Escobar, A. 1994. Welcome to cyberia—notes on the anthropology of cyberculture. *Current Anthropology* 35(3): 211–321.
- Etchevers-Goijberg, N. (2005). Ruta etnográfica para la comprensión de la comunicación on-line”. Disponible: <http://www.cibersociedad.net/archivo/articulo.php?art=204>. Consulta: Octubre 10, 2012.
- Faura, R. (2000). Cibercultura, realidad o invencion?. Disponible en: [http://www.naya.org.ar/congreso2000/ponencias/Ricard\\_Faura.htm](http://www.naya.org.ar/congreso2000/ponencias/Ricard_Faura.htm). Consulta: Octubre 10, 2012.
- Festinger, L. (1978). *A theory of cognitive dissonance*. Evanston, Ill.: Row and Peterson.
- Fishbein, M. (1965). A consideration of beliefs, attitudes and their relationships. En I. D. Steiner y M.

- Fishbein (eds). (1965). *Studies in social power*. Ann Arbor: Institute for Social Research.
- Flanders, N. A. (1977). *Análisis de la interacción didáctica*. Salamanca: Anaya.
- Flores, F. K. M. (2012). Metodología PACIE en los ambientes virtuales de aprendizaje para el logro de un aprendizaje colaborativo. Disponible: <http://dialnet.unirioja.es/servlet/articulo?codigo=4156135>. Consulta: Septiembre, 2013.
- Frey, J., y Fontana, A. (1993). The group interview in social research. In D. Morgan (Ed.), *Successful Focus Groups: Advancing the State of the Art* (pp. 20-34). Newbury Park, CA: Sage.
- Fruchter, B. (1954). *Introduction to Factor Analysis*. New York: Van Nostrand. Gabaldón-Mejía, N. (1973). *Algunos conceptos de muestreo*. Caracas: UCV.
- Gabaldon Mejia, N. (1973). *Algunos conceptos de muestreo*. Caracas: UCV.
- Garrido, A. (2003). El aprendizaje como identidad de participación en la práctica de una comunidad virtual. Disponible: [www.uoc.edu/in3/dt/20088/20088.pdf](http://www.uoc.edu/in3/dt/20088/20088.pdf). Consulta: octubre 22, 2012.
- Garner, W. R., y Creelman, C. D. (1967). Problems and methods of psychological scaling. En H. Helson y W. Bevan. *Contemporary approaches to psychology*. Princeton: Van Nostrand Co.
- Garrison, D. R., Anderson, T. (2010). *El E-Learning en siglo XXI. Investigación y práctica*. Barcelona: Octaedro, S. L.
- Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: interaction is not enough. *American Journal of Distance Education*, 19(3), 133–148.
- Garrison, D. R., Anderson, T, & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education* 2: 87–105.
- Garrison, D. R., & Archer, W. (2000). *A transactional perspective on teaching and learning: A framework for adult and higher education*. Oxford, UK: Pergamon.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American*

*Psychologist*, 18, 519-521.

Glaser, B. G., y Strauss, A. M. (1967). *The discovery of grounded theory*. Chicago: Aldine Publishing Co.

Glitz, B. (1998). *Focus groups for libraries and librarians*. New York: Forbes Custom Publishing.

Goetz, J. P., y LeCompte, M. D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.

González, G. T., y Cano, A. A. (2007). Introducción al análisis de datos en investigación cualitativa: Tipos de análisis y proceso de codificación. Disponible en:

[http://www.nureinvestigacion.es/ficheros\\_administrador/f\\_metodologica/analisisdatoscodif45.pdf](http://www.nureinvestigacion.es/ficheros_administrador/f_metodologica/analisisdatoscodif45.pdf).

Consulta realizada: Agosto 30, 2012.

Gorsuch, R. L. (1966). The general factor in the test anxiety questionnaire. *Psychological Reports*, N° 19, p.

Green, B. F. (1954). Attitude measurement. En G. Linsey (ed). *Handbook of Social Psychology*. Cambridge, Mass.: Addison- Wesley.

Gronlund, N. E. (1976). *Elaboración de test de aprovechamiento*. México: Trillas.

Guasch O. (1997). Observación Participante. *Cuadernos Metodológicos*, N° 20, Madrid: CIS.

Guba, E. G., y Lincoln, Y. (1992). *Fourth generation evaluation*. New Delhi: Sage Publications.

Guba, E. G., y Lincoln, Y. (1982). *Effective evaluation*. San Francisco: Jossey-Bass Publishers.

Gunawardena, C. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferencing. *International Journal of Educational Telecommunications*, 1(2-3), 147-166.

Guertin, W. H., and Bailey, J. P. (1970). *Introduction to modern factor analysis*. Ann Arbor, Michigan: Edwards Brothers, Inc.

- Guilford, J. P. (1954). *Psychometric Methods*. New York: McGraw- Hill Book Co.
- Guttman, L. (1947). Some necessary conditions for common factor analysis. *Psychometric*, N° 19, 149-161.
- Guttman, L. (1954). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- Hakken, D. (1999). *Cyborgs@Cyberspace?: an ethnographer looks to the future*. New York: Routledge.
- Hannerz, U. (1986). *Exploración de la ciudad*. México: Fondo de Cultura Económica.
- Harman, H. (1976). *Modern Factor Analysis*. (2da ed). Chicago: University of Chicago Press.
- Havelock, R. G., y Huberman, A. M. (1980). *Innovación y problemas de la educación*. París: UNESCO.
- Helmstadter, G. C. (1964). *Principles of psychological measurement*. New York: Appleton.
- Hendrichson, A. E., y White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*. N° 17, 65-70.
- Henerson, M. E., Lyons-Morris, L., and Taylor, C. (1978). *How to measure attitude*. Beverly Hills, Ca.: Sage Publications, Inc.
- Hernández, S. R., Fernández, C. C., y Baptista, L. P. (1991). *Metodología de la investigación*. México: McGraw-Hill.
- Hine C. (2004). *Etnografía virtual*. Barcelona: Editorial UOC. Colección Nuevas Tecnologías y Sociedad, pp. 210.
- Hine, C. (2000). *Virtual Ethnography*. London: Sage.
- Holzinger, K. (1936). *Factor Analysis*. Chicago: University of Chicago Press.
- Holzinger, K., y Harman, H. H. (1941). *Factor Analysis*. Chicago: University of Chicago Press.

- Hotelling, H. (1931). The generalization of student ratio. *Ann Math Statist*, N° 2, 360-378.
- Hotelling, H. (1936). Simplified calculation of principal components. *Psychometrika*, 1, 27-35.
- Hovland, C. I., Lumsdain, A. A., and Sheffield. (1949). *Experiment on mass communication*. Princeton: Princeton University Press.
- Hoyt, C. (1941). Test reliability obtained by analysis of variance. *Psychometrika*, 6, 153-160.
- Husserl, H. (1962). *Ideas relativas a una fenomenología pura y una filosofía fenomenológica*. México: FCE.
- Hurtado de Mendoza, M A. (1980). *Pruebas de rendimiento académico y objetivos de la instrucción*. México: Diana.
- Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factory analysis. *Psychometrika*, N° 23, 187-200.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*. N° 20, 141- 151.
- Kaiser, H. F. (1970). A Second Generation Little Jiffy. *Psychometrika*, N° 35, 401-417.
- Kaiser, H. F., y Rice, J. (1974). Little Jiffy Mark IV. *Educational and Psychological Measurement*, 34, 111-117.
- Kaplan, B., y Duchon, D. (1988). Combining qualitative and quantitative methods in information Systems research: A case study, *MIS Quarterly*, december 1988, pp.571-586.
- Katz, D., y Scottand, E. (1976). *A preliminary statement to a theory of attitude structure and change*. New York: McGraw-Hill.
- Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health and Illness* 16(1), 103–121.

Krueger, R., y Casey, M. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Newbury Park, CA: Sage.

Keller, F. S. (1968). Good-bye teacher. *Journal of Applied Behavioral Analysis*, 1, 79-89.

Kerlinger, F., y Lee, H. B. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales*. México: McGraw-Hill.

Kidder, L. H., y Campbell, D. T. (1976). El examen directo de las actitudes sociales. En G. F. Summers. *Medición de actitudes*. México: Trillas.

Kish, L. (1965). *Survey sampling*. New York: Wiley and Sons.

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health and Illness* 16(1), 103–121.

Kluasmeier, H. J., y Goodwin, W. (1966). *Learning and human abilities*. (2da. ed). New York: Harper and Row.

Krech, D., Crutchfield, R. S., y Balachey, E. L. (1962). *Theory and problems of social psychology*. New York: McGraw-Hill.

Kuder, G. F., y Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2, 151-160.

Kuhn, T. (1978). *La estructura de las relaciones científicas*. Mexico: FCE.

Lafourcade, P. (1969). *Evaluación de los aprendizajes*. Buenos Aires: Kapelusz.

Lemus, L. A. (1974). *Evaluación del rendimiento escolar*. Buenos Aires: Kapelusz.

Liamputtong, P. (2010). Focus group methodology: Introduction and history. Disponible: [http://www.sagepub.com/upm-data/39360\\_978\\_1\\_84787\\_909\\_7.pdf](http://www.sagepub.com/upm-data/39360_978_1_84787_909_7.pdf). Consulta realizada en: Agosto 07, 2010.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology* N° 140.

Lindon, A. (2005). La construcción social de los paisajes invisibles y del miedo. III Seminario Internacional sobre el Paisaje: Paisajes Incógnitos, Territorios Ocultos: Las Geografías de la Invisibilidad, Observatorio del Paisaje. Disponible: <https://es.scribd.com/doc/160277122/Lindon-Paisajes-Invisibles-y-Del-Miedo>. Consulta: octubre 18, 2012.

Lindvall, C. M. (1967). *Measuring pupil achievement and aptitude*. New York: Hartcourt, Brace and World, Inc.

Loevinger, J. A. (1947). A systematic approach to the construction and evaluation of test of ability. *Psychological Monographs*, 61(4).

Madriz, E. (2003). Focus groups in feminist research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd ed). Thousand Oaks, CA: Sage Publications.

Magnusson, D. (1982). *Teoría de los tests*. México: Trillas.

Manning, S. A., y Rosenstock, E. H. (1971). *Elaboración de escalas de actitudes y psicofísica clásica*. México: Trillas.

Malinowski, B. (2007). Etnografía virtual. Disponible: <http://lastrobiand.blogspot.com/2007/05/etnografia-virtual.html>. Consulta: Octubre: 15, 2012.

Martínez, M. (1989). *Comportamiento humano: Nuevos de investigación*. Mexico: Trillas.

Martínez, M. M. (2004). *La investigación cualitativa etnográfica en educación*. México: Trillas.

Martínez, M. (1989). *Comportamiento humano. Nuevos métodos de investigación*. México: Trillas.

Mayans, J. (2002). Género Chat. O Cómo la etnografía puso un pie en el ciberespacio, Barcelona: Gedisa.

Más información en <http://cibersociedad.rediris.es/mayans/genchat.php>.

McGuire, W. J. (1969). The nature of attitudes and attitude change. En G. Lindzey and E. Aronson

(eds). *Handbook of social psychology*. Los Angeles: Addison-Wesley.

Mehrabian, A. (1969). Male and female scales of the tendency to achieve. *Educational and Psychological Measurement*, 28, 444-493- 502.

Mendez-Castellanos, H. (1985). *Proyecto Venezuela*. Caracas: FUNDACREDESA.

Merton, R. K. (1987). The focused interview and focus groups. *Public Opinion Quarterly*, 51, 550-566.

Miles, M. B., y Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills (Ca): Sage Publications, Inc.

Ministerio de Educación. (1986). *Reglamento general de la ley orgánica de educación*. Caracas: Romar.

Ministerio de Educación (1980). *Ley orgánica de educación*. Caracas: Romar.

Ministerio de Educación. (1970). Características socio-económicas, status social y actitudes de los docentes venezolanos. Departamento de Investigaciones Educativas. Caracas: Autor.

Montero, A. (2012). Significados cognitivos, sociales y afectivos de las comunidades virtuales de aprendizaje: Una visión intersubjetiva desde la UPEL-IPB. Tesis de grado no publicada presentada a la ilustre Universidad Pedagógica Experimental Libertador para optar al grado de Doctor en Educación.

Montero, M. (2009). *Grupos focales*. Caracas: AVEPSO.

Montgomery, D. C. (1984). *Design and Analysis of Experiments*. New York: Wiley and Sons.

Morgan, D. L. (2002). Focus group interviewing. In J. F. Gubrium y J. A. Holstein (eds.), *Handbook of interviewing research: Context & method* (pp. 141-159). Thousand Oaks, CA: Sage.

Mudarra, M. A. (1978). *Historia de la legislación escolar contemporánea en Venezuela*. Caracas: Publicaciones Mudbell.

Neuhaus,, J. O., y Wrigley, V. (1954). The quartimax method: An analytical approach to orthogonal simple structure. *British Journal of Statistical Psychology*, N° 7, 81-91.

Neve, E. (2006). *Exploración del espacios y lugares digitales a través de la observación flotante. Una propuesta metodológica*. Disponible:  
[http://www.uned.es/etnovirtual/GT\\_OCS\\_etnografias%20digital\\_comunicaciones.pdf](http://www.uned.es/etnovirtual/GT_OCS_etnografias%20digital_comunicaciones.pdf). Consulta:  
Octubre 23, 2012

Nunnally, J. C. (1987). *Teoría Psicométrica*. México: Trillas.

Nunnally, J. C., y Berstein, I. J. (1995). *Teoría Psicométrica*. (3ra ed). México: McGraw-Hill.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw- Hill.

Nunnally, J. C. (1972). *Educational measurement and evaluation*. New York: McGraw-Hill.

Olivares, M. (1984). *Innovaciones en planeamiento de la educación en América Latina*. Caracas: Publicaciones de

OPSU-CENAMEC. (1987). *Diagnóstico del nivel de conocimiento en biología, matemática, química, física, ciencias de la tierra y lenguaje en estudiantes del ciclo básico de educación media*. Caracas: Autor.

Osgood, C. E., Suci, G. J., y Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana University of Illinois Press.

Patton, M. (1990). *Quality evaluation methods*. Beverly Hills. Sage.

Pettonnet, C. (1982). "L'Observation Flotante: L'exemple d'un cimetière parisien", *L'Homme*, Vol. 22, Num 4, p. 37-47.

Perera, V. H. (2007). *Estudio de la interacción didáctica en elearning*. Tesis doctoral presentada a la ilustre Universidad de Sevilla. Disponible: <http://hera.ugr.es/tesisugr/1850436x.pdf> Consulta:  
Octubre 03/2012.

- Piaget, J. (1978). *Psicología de la inteligencia*. Madrid: Critica.
- Popham, W. J. (1978). *Problemas y técnicas de la evaluación educativa*. Madrid: Anaya.
- Popham, W. J. (1980). *Criterion-referenced measurement*. Englewood Cliffs, N. J.: Prentice-Hall.
- Popham, W. J., y Husek, T. R. (1969). Implications of criterion referenced measurement. *Journal of Educational Measurement*, 6,1-9.
- Provus, M. M. (1973). Evaluation of ongoing programs in the public school system. En B. R. Northen and J. R. Sanders (eds). *Educational evaluation: Theory and Practice*. Northington, Ohio: Charles A. Jones Publishing.
- Refez, L. (1972). *Social psychology*. Blenviews, Ill.: Scott Foresman.
- Reguillo, R. (1998). De la pasión metodológica o la (paradójica) posibilidad de la investigación. En, Mejía A. R. y Sandoval. S. A. (1998), (coords.), *Tras las vetas de la investigación cualitativa. Perspectivas y acercamientos desde la práctica*, ITESO, México, pp. 17-38.
- RAE. (2001). *Diccionario de la lengua española*. (22ª ed.). Madrid: Espasa.
- Rennekamp, R. A., y Nall, M. A. (2001). *Using Focus Groups in Program Development and Evaluation*. Disponible en: <http://www.ca.uky.edu/agpsd/focus.pdf>. Consulta: Agosto 9, 2012.
- Rodriguez, G. P. (1997). *El recurso informático en el procesamiento de datos cualitativos en ciencias sociales*. Disponible en: [http://www.naya.org.ar/analisis\\_cualitativo/vcaaspc.htm](http://www.naya.org.ar/analisis_cualitativo/vcaaspc.htm). Consulta realizada: Octubre 12, 2010.
- Rodríguez G. G., Gil J. y García J. E. (1996). *Metodología de la investigación cualitativa*. Málaga (España): Ediciones Aljibe
- Rodríguez, H. (2012). *Atribuciones de éxito y fracaso en las iniciativas institucionales de e-Learning*. Tesis doctoral no publicada presentada a la ilustre Universidad Nacional Experimental Libertador

como un requisito parcial para optar al grado de Doctor en Educación. Barquisimeto (Venezuela), Abril 2012).

Rosenshine, B. V. (1980). How time is spent in elementary classrooms. In C. Denham y A. Leiberman (eds.). *Time to learn*, Washing-ton, D.C.:Department of Health, Education, and Welfare, National Institute of Education.

Rourke, L., Anderson, T. Garrison, D. R., & Archer, W. (1999). Assessing Social Presence in Asynchronous, Text-Based Computer Conferencing. *Journal of Distance Education*, 14, (3), 51-70.

Ruiz Bolívar, C. (2006). *Como llegar a ser un tutor competente*. Caracas: Santillana/UPEL.

Ruiz, B. C. (1994). Escala de satisfacción en la lectura. No publicada. Barquisimeto, Mimeo.

Ruiz, B. C. (1991). Desarrollo de una escala de actitud hacia la innovación educativa. *Investigación y Posgrado*, 6, 1,49-91.

Ruiz, B. C. (1990). *Desarrollo de una escala de actitud hacia la innovación educativa*. Trabajo para ascender a la categoría de profesor asociado en la Universidad Pedagógica Experimental Libertador. Caracas. Mimeo.

Ruiz, B. C. (1989). Escala AC-2000: Evaluación del autoconcepto en estudiantes de educación básica. *RevistaParadigma*.Vol X, 2, 201- 217.

Ruiz, B. C. (1988). Uso y Abuso de la escala Likert en la investigación psicoeducativa. *Investigación y Posgrado*, 3(1) 93-112.

Ruiz, B. C. (1983). Efectos del programa Enriquecimiento Instrumental sobre los factores cognoscitivos y no- cognoscitivos en sujetos de diferentes estratos socioeconómicos. Puerto Ordaz. Publicaciones de la Corporación Venezolana de Guayana. Mimeo.

Ruiz Olabuenaga, J.I.: (1996). *Metodología de la investigación cualitativa*. Bilbao. Universidad de Deusto.

Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9,99-103.

Ryans, D. G. (1951). The results of internal consistency and external validation procedures applied in the analysis of test items measuring professional information. *Education and Psychological Measurement*, 11,549-560.

Sachs, A. G. (1970). *Medición y evaluación*. Barcelona (España): Herder.

Salom de Bustamante, C. (1981). *Dos medidas de necesidades de logro: Una comparación*. Mérida: Laboratorio de Psicología, ULA. Publicación 19.

SALVAT. (1968). *Diccionario enciclopédico Salvat* (13a ed). Madrid: Autor.

Sánchez de Hurtado, J. (1994). Efecto de la retroalimentación del docente en algunas variables psicoafectivas del alumno y su transferencia al aprendizaje de estudiantes de planteles públicos y privados. Tesis doctoral no publicada, presentada a la Universidad Nacional Experimental Simón Rodríguez.

Sánchez, M. (1991). *Desarrollo de habilidades del pensamiento. Procesos básicos del pensamiento*. México: Trillas.

Sánchez-Upegui, A. A. (2010). Nuevos modos de interacción educativa: análisis lingüístico de un foro virtual. *Revista Educación y Educadores*, vol. 12, No. 2.

Scriven, M. (1967). The methodology of evaluation. En R. E. Stake (ed). *Curriculum and evaluation*. American Educational Research Association Monograph series en evaluation N° 1. Chicago: Rand McNally.

Seidman, I. (1998). *Interviewing as Qualitative Research*. (2ª ed.). New York: Teachers College Press.

Seijas, Z. F. (1981). *Investigación por muestreo*. Caracas: Publicaciones UCV.

Shaw, M. E., y Wright, J. M. (1967). *Scales for the measurement of attitudes*. New York: McGraw-Hill.

Shea, P., Li, C., Swan K., Pickett, A. (2005) Developing Learning Community in Online Asynchronous College Courses: The Role of Teaching Presence. *Journal of Asynchronous Learning Networks*, 9(4). Retrieved May 1, 2008 from:

<http://www.sciencedirect.com/science/article/pii/S1096751606000364>

Sierra-Bravo, R. (1987). *Tesis doctorales y trabajos de investigación científica*. Madrid: Paraninfo.

Silva, M., De Sanctis, B., y Saarco, L. (1982). *Proceso de la construcción y análisis de los resultados de la Prueba de Aptitud Académica: Reporte técnico*. Caracas: CNU-OPSU.

Soto, J. (1999). Psicología de lo borroso. XXVII Congreso Interamericano de Psicología, Caracas, Venezuela. Disponible: <http://www.uam.mx/difusion/revista/sep2000/soto.html>. Consulta: Octubre 20, 2012.

Spearman, C. General Intelligence objectively determined and measured. *American Journal of Psychology*, N° 15, 201-293.

Spradley, J. P. (1980). *Participant observation*. New York: Rinehart & Winston.

Stake, R. E. (1967). The countenance of educational evaluation. *Teacher College Record*, 68,523-540.

Stanley, J. C. (1974). *Measurement in today's schools*. Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Stevens, S. S. (1951). Mathematics, measurement, and Psychophysics. En S. S. Stevens (ed). *Handbook of experimental psychology*. New York: Wiley and Sons.

Strathern, M. (1996). Cutting the Network. *Journal of the Royal Anthropological Institute*, 2, pp. 517-535.

Strauss, A., y Corbin, J. (2002). *Bases de la investigación cualitativa*. Antioquia: Imprenta dela

Universidad de

Strauss, A., y Corbin, J. (1998). *Basics of qualitative research. Techniques and procedures for developing grounded theory*. New York: Sage Publications Inc.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., Provus, M. M. *Educational evaluation and decision making in education*. Itasca, Ill.: Peacock.

Summers, G. F. (1976). *Medición de actitudes*. México: Trillas.

Summers, G. F., y Hammonds, A. D. (1969). Toward a paradigm for respondent bias in survey research. *Sociological Quarterly*, 10,113-121.

Swan, K., Garrison, D. R., y Richardson, J. C. (2009). A constructivist approach to online learning: the Community of Inquiry framework. In Payne, C. R. (Ed.) *Information Technology and Constructivism in Higher Education: Progressive Learning Frameworks*. Hershey, PA: IGI Global, 43-57.

Tashakkori, A., and Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research*. London: Sage.

Tartarini, E. (1967). *Evaluación escolar y elementos de estadística aplicada*. Santiago: Editorial Universitaria.

Taylor, S. J., y Bogdan, R. (1996). *Introducción a los métodos cualitativos de investigación*. México: Paidós.

Téllez-Infantes, A. (2002). *Nuevas etnografías y ciberespacio: reformulaciones metodológicas*. Disponible: <http://www.cibersociedad.net/congreso/comms/g10tellez.htm>. Consulta: Octubre 10, 2012.

Thorndike, R. L. (1989). *Psicometría aplicada*. México: Limusa.

Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological Review*, 36,222-241.

- Thurstone, L. L. (1931). Multiple Factor Analysis. *Psychological Review*.38,406-427.
- Thurstone, L. L. (1947). *Multiple Factor Analysis*.Chicago: University of Chicago Press.
- Toro, A. F. (1992). *Cuestionario de motivación para el trabajo: Manual de instrucciones* (2da. ed). Medellín, Colombia: Cincel, Ltda.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35,492-501.
- Vigotsky, J. (1981). *Pensamiento y Lenguaje*. Buenos Aires: La Pléyade.
- Tyler, R. (1934). *Modelos de evaluación: Pensar en evaluación de servicios educativos y humanos*. Norwell, Mass.: Kluwer Academic Publisher
- Vigotsky, J. (1981). *El desarrollo de los procesos psicológicos superiores*. Buenos Aires: Grijalbo.
- Volkova, B. D. (1953). Some characteristics of conditioned reflex formation to verbal stimuli in children. *Schenov Journal pf Psychology*, 39, 540-548.
- Webb, E. J., y Salancik, J. R. (1976). Suplemento al informe sobre si mismo en la investigación de actitudes. En G. F. Summers. *Medición de actitudes*.México: Trillas.
- Whittaker, J. O. (1979). *La Psicología social en el mundo de hoy*.México: Trillas.
- Wilkinson, S. (2004). Focus groups: A feminist method. In S.N. Hesse-Biber y M. L. Yaiser (eds.), *Feminist perspectives on social research* (pp. 271–295). New York: Oxford University Press.
- Winer, J. B. (1971). *Statistical Principles in Experimental Design*(2<sup>nd</sup> ed). New York: McGraw-Hill.
- Wood, D., y Bruner, J., y Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89-100.
- Woods, P. (1987). *La escuela por dentro: la etnografía en la investigación educativa*. Barcelona (España): Paidós/MEC.