



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE PSICOLOGIA
División de Estudios de Posgrado

2
20

UNA CONTRIBUCION TECNOLOGICA AL DIAG-
NOSTICO PSICOEDUCATIVO, BASADA EN LA
SIMULACION NEUROCOMPUTACIONAL

T E S I S

QUE PARA OBTENER EL GRADO DE:
DOCTOR EN PSICOLOGIA GENERAL
EXPERIMENTAL

P R E S E N T A:

MIGUEL LOPEZ OLIVAS

Directora de Tesis: Dra. Isabel Reyes Lagunes

COMITE:

Dr. Rolando Díaz Loving
Dr. José de Jesús González Núñez
Dr. Serafín Mercado Domenech
Dr. Javier Nieto Gutiérrez

SINODALES:

Dra. Graciela Rodríguez Ortega
Dr. Juan José Sánchez Sosa



México, D. F.

Noviembre de 1996

TESIS CON
FALLA DE ORIGEN

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

Expreso a la Dra. Isabel Reyes Lagunes mi más profunda gratitud, aunada a un reconocimiento por demás sincero, porque al dirigirme en la realización de esta tesis me benefició con su talento, dedicación y, sobre todo, con su amistad.

De igual manera, hago patente un amplio agradecimiento a mi Comité de Tesis, integrado por el Dr. Rolando Díaz-Loving, el Dr. José de Jesús González Núñez, el Dr. Serafín Mercado Domenech y el Dr. Javier Nieto Gutiérrez. Y a mis Sinodales, la Dra. Graciela Rodríguez Ortega y el Dr. Juan José Sánchez Sosa.

Su participación me representó un altísimo honor, por tratarse de universitarios muy prestigiados y queridos en nuestra comunidad académica. Gracias a su consejo, estímulo y apoyo pude avanzar en este trabajo y llegar a la meta final.

Deseo también dejar constancia de mi mayor estimación a un psicólogo de corazón, el Dr. José Manuel Álvarez Manilla, quien en forma amigable y experta me hizo compartir su visión constructiva y futurista de la innovación tecnológica en la evaluación educativa.

Agradezco al Dr. Roberto Castañón Romo por distinguirme con su confianza y permitirme participar en dos proyectos de investigación y aplicación, que a la vez que enriquecieron mi formación, me dieron la oportunidad de contribuir en la siempre significativa tarea de impulsar la vida académica de nuestra querida Universidad.

ÍNDICE

Resumen.	1
Introducción.	5
CAPÍTULO I: La perspectiva psicométrica en tareas de reconocimiento de patrones.	9
a) Introducción.	9
b) Principios y componentes de los modelos IRT.	14
c) Aspectos procedimentales de los modelos IRT.	23
d) Componentes de una estrategia metodológica para la evaluación integral del desempeño escolar.	28
CAPÍTULO II: La perspectiva neurocomputacional en tareas de reconocimiento de patrones.	36
a) Introducción.	36
b) Algunos antecedentes del campo de la neurocomputación, relevantes para la psicología.	54
c) Arquitectura y operación del prototipo en neurocomputador.	64
CAPÍTULO III: El problema de investigación y diseño del estudio.	72
CAPÍTULO IV: Método	84
1. Sujetos.	84
2. Materiales e instrumentos:	84
A. en la fase psicométrica.	84
B. en la fase neurocomputacional.	84
3. Procedimiento:	85
A. en la fase psicométrica.	85
B. en la fase neurocomputacional.	90
CAPÍTULO V: Resultados.	93
Introducción.	93
A.1. Resultados del análisis psicométrico:	
A.1.1. Estructura de los grupos y de los instrumentos de evaluación.	93
A.1.2. Resultados totales de los grupos de comparación.	95
A.1.3. Análisis para probar la equivalencia de grupos y versiones	100

de examen.	
A.1.4. Sobre la homogeneidad de los grupos: resultados por nivel de desempeño.	102
A.1.5. Análisis para probar la equivalencia de grupos y versiones de examen por niveles de desempeño.	107
A.2. Análisis psicométrico	114
B. Resultados del análisis neurocomputacional.	131
Introducción.	131
B.1. Obtención de resultados en el procesamiento neurocomputacional.	133
B.2. Resultados del procesamiento neurocomputacional.	141
CAPÍTULO VI: Discusión.	146
A. Discusión de los resultados en la porción psicométrica:	
A.1. Discusión de los resultados globales.	146
A.2. Discusión de los perfiles de desempeño.	157
B. Discusión de los resultados en la porción neurocomputacional.	160
CAPÍTULO VII: Sumario y conclusiones.	163
Referencias bibliográficas.	177
Anexos:	185
1. Resultados de la evaluación de los sujetos.	
2. Resultados del análisis psicométrico de los reactivos.	

RESUMEN

La presente investigación se enmarcó teórica y metodológicamente en la evaluación del desempeño escolar de estudiantes del nivel educativo de enseñanza media, atendiendo al dominio de conocimientos y habilidades académicas (comprensión de la lectura, solución de problemas aritméticos y matemáticos, manejo conceptual, aplicación de reglas y principios), en la realización de actividades diagnóstico-prescriptivas del aprendizaje.

El análisis de las respuestas en un examen estándar fue la fuente para comparar los patrones asociados al número de aciertos, permitiendo identificar perfiles diferenciales del desempeño a la luz de dos parámetros: habilidad de los sujetos y dificultad de los ítems, y con el apoyo de tres indicadores: contenido programático, tipos de habilidad y de conocimiento.

El problema de investigación que alimentó este trabajo, surgió del interés por someter a prueba la hipótesis de que el reconocimiento de patrones de desempeño, inherente al diagnóstico psicoeducativo, puede hacerse confiable y eficientemente si se aplican técnicas y procedimientos propios de dos campos metodológicos: psicométrico e Inteligencia Artificial y si se dispone de información suficiente y de buena calidad.

La condición expresada indicó la conveniencia de diseñar un estudio de dos etapas. La primera orientada a medir el desempeño de sujetos con distintos niveles de habilidad, para efectuar análisis psicométricos de las respuestas y determinar los perfiles diferenciales; la segunda tendría como propósito diseñar, desarrollar y probar un sistema que operara como un experto automatizado al efectuar dicho reconocimiento.

El sistema tomó forma en un prototipo neurocomputacional, con capacidad para aprender a identificar los patrones de interés confiablemente y con velocidad considerable.

En este contexto, para la primera fase se revisaron principios teóricos y psicométricos relevantes, derivados de la *Teoría de la Respuesta de Ítems* (abreviada en inglés *IRT*), que han sido reportados en la literatura especializada, acerca de la construcción de instrumentos para medir el desempeño. Con su utilización se buscó impactar positivamente la calidad del proceso y los productos de la evaluación planeada.

El aporte psicométrico se aplicó a explorar, en un nivel de *grano fino*, los rasgos diferenciales del desempeño académico subyacentes a tres patrones de habilidad: alta, media, baja, en términos de la disponibilidad de conocimientos y habilidades, en ocho dominios temáticos que están considerados en la evaluación para ingresar a bachillerato (CENEVAL, 1994).

La segunda etapa del estudio, apoyada en tecnología de la *simulación neurocomputacional*, tomó los resultados de la evaluación de un grupo de sujetos (fase psicométrica) para entrenar el prototipo. Con esa base se hizo la primera prueba de precisión del reconocimiento automatizado de patrones. Después, se procedió a probar el neurocomputador con un número menor de sujetos no incluidos en la fase de entrenamiento, pero que formaban parte del grupo original y cuyas características psicométricas eran conocidas. Por último, se evaluó si el sistema era útil para identificar patrones de sujetos de otros grupos, abarcando una población total de 1500.

De esta manera, para realizar el reconocimiento de patrones y la categorización de los estudiantes, se tuvo a manera de "juez" primero un entorno parcialmente diferente y luego uno distinto del que se empleó para el entrenamiento, logrando un nivel de confiabilidad óptimo.

En la fase de entrenamiento del prototipo, utilizando una computadora PC 486 DX de 66 Mhz y memoria RAM de 22 megabytes, el procesamiento requirió alrededor de 36 minutos para efectuar un promedio de 1,000 lecturas de los datos completos, con un total de 249,000 ciclos de ajuste y error = cero.

Posteriormente, se procedió al agrupamiento de los datos de los sujetos de los cinco grupos en una sola matriz de prueba (N = 1500), entre los que se encontraban los 300 utilizados inicialmente. Nuevamente, el error en el reconocimiento de los patrones de desempeño fue cero, requiriendo para esta tarea alrededor de un minuto.

Por último, en una muestra de estudiantes de preparatoria se aplicó una variante de la estructura del prototipo para identificar y evaluar patrones de respuesta, en una *prueba de habilidades intelectuales* (CISE, 1995) de tipo piagetiano, que distinguió cuatro categorías: concreto inicial, concreto avanzado, formal inicial y formal avanzado. Esta aplicación experimental tuvo por meta probar la operación neurocomputacional en un contexto diferente al que sirvió de base general a esta investigación. De nuevo, el entrenamiento de la red requirió de minutos; el reconocimiento de patrones se efectuó en segundos.

Los resultados de esta investigación aportaron evidencia substancial, en apoyo a la utilización confiable de la simulación neurocomputacional, en tareas que requieren del reconocimiento de patrones, como el diagnóstico psicoeducativo; y prueban la capacidad de alta generalización propia de dicha innovación tecnológica, ya que su uso está ligado, dicho simplemente, al reconocimiento preciso de patrones y de secuencias de acciones, indistintamente del objeto de estudio que se seleccione para su utilización.

ABSTRACT

This research was based theoretically and methodically on scholastic achievement evaluation of high school students, measuring knowledge handle and academic skills on the realization of diagnostic-prescriptive tasks. The goal was to identify the differential patterns with two parameters: person ability and test difficulty, using psychometry supported by the Item Response Theory and the neurocomputational simulation risen from Artificial Intelligence.

The analysis was made upon the answers of a 1,500 individuals random sample, divided in 5 independent groups of 300 students, to measure achievement: high, medium and low in 8 knowledge domains and three different skills.

The outcomes showed experimental evidence supporting automatic recognition of differential achievement patterns and it was proved that neurocomputational simulation represents a technological innovation with highly generalization in educational evaluation.

Key words: differential scholastic achievement, IRT psychometric models, neurocomputational simulation.

TESIS

COMPLETA

INTRODUCCION

El presente trabajo tuvo su origen en las líneas de investigación orientadas en los marcos de la psicología cognoscitiva, de la psicología instruccional y de la psicometría cognoscitiva, realizadas desde hace varios años en las áreas educativa y experimental del posgrado de la Facultad de Psicología de la UNAM. Su objetivo central ha sido explorar con profundidad diversos aspectos del aprendizaje universitario de estudiantes con distintos niveles de desempeño académico, como base para aportar estrategias, materiales, programas y modelos teóricos y metodológicos que puedan contribuir a su enriquecimiento.

De esta manera, se ha indagado la significatividad de la relación entre diferentes componentes del aprendizaje complejo, como: habilidades para la lectura y éxito escolar; estrategias de aprendizaje y ejecución de tareas de alta demanda cognoscitiva; componentes metacurriculares y niveles de desempeño, entre otros. Las variables seleccionadas han recibido sustento teórico y empírico de estudios realizados por Johnston (1989), Pozo (1990), De Vega (1989), Bransford (1979), Gagnè (1995) y Anderson (1983), además de trabajos propios realizados en el posgrado de la Facultad de Psicología de la UNAM: Castañeda y López (1988a y 1988 b; 1990 y 1991), Castañeda, López y Romero (1987), Castañeda, López, Gómez, Cabrera y Orozco (1989), López y Castañeda (1990, 1992), López, Castañeda y Gómez (1989) y López, Castañeda, Pineda y Orduña (1992).

En estos estudios ha estado presente el interés de comparar los efectos de variables relacionadas con el desempeño escolar, buscando identificar qué es lo que hace diferentes a los estudiantes académicamente exitosos de los que no lo son. Por ejemplo, explorar cuál es la diferencia en el desempeño entre alumnos que dan respuestas distintas en un inventario acerca de la orientación al logro, autoestima e internalidad; qué aspectos son distintivos de los resultados en los cursos, en alumnos poco hábiles para utilizar estrategias de solución de problemas, etc.

El énfasis en la realización de estos estudios ha estado puesto, a veces no muy propositivamente, en la identificación de características o rasgos asociados a un estado particular del alumno, que definen variables que se vinculan con patrones observables y medibles. Se asume que estas variables subyacen a las descripciones de los sujetos cuando se les rotula como exitosos o deficientes, implicando la existencia de patrones más o menos típicos. De alguna forma, el modelo de investigación de novatos-expertos a que se refieren Glaser y Bassok (1989) y a cómo pasar de un estado al otro, han estado en el foco de interés de las investigaciones realizadas.

Las crecientes capacidades computacionales también han impactado los avances logrados en dicha investigación. De esta forma, se cuenta con prototipos que conjuntan aplicaciones originales. Por ejemplo, se desarrolló un sistema experto llamado *Thor-Ombolo* (Castañeda y López, 1991), escrito en lenguaje LISP para apoyar las tareas de diagnóstico y prescripción de los riesgos en el aprendizaje académico, conjuntando la recomendación de medidas para superar los problemas detectados.

Otros prototipos interesantes por su operación y resultados han sido el diseño y la construcción de varios diagnosticadores generados en el entorno de programación de bases de datos. El prototipo *Asesor* (López y Castañeda 1990) y los sistemas *SIETE* y *SEPA* (López, Castañeda y Gómez, 1989), son muestras de ellos.

Una línea de estudio que en particular aportó elementos importantes para este trabajo, buscó contestar la pregunta: ¿cambia la representación del conocimiento después de la lectura de un texto? Para ofrecer respuestas plausibles a esta interrogante se realizaron dos estudios, uno a nivel de la *microestructura del texto* (Castañeda, 1993) y el otro a nivel del texto como *unidad* (López, Castañeda, Pineda y Orduña, 1992), que involucraron tareas experimentales de lectura y elaboración de mapas semánticos, como formas para analizar la representación del texto elaborada por el lector.

El punto de interés común en las investigaciones fue explorar cómo se representaban el contenido sujetos con distintos niveles de conocimiento, acerca del tópico del texto que fue seleccionado, asumiendo que se trataba de lectores hábiles (estudiantes del nivel profesional y de posgrado). De esta forma se intentó identificar los perfiles de cuatro categorías: (a) sin y (b) con *conocimiento previo*, (c) antes y (d) *después de leer* el texto.

Una preocupación inherente al planteamiento primario de identificar perfiles diferenciales, fue explorar las perspectivas tecnológicas disponibles para procesar y analizar rápida y confiablemente grandes volúmenes de información que se generaban al hacer estudios e intervenciones en los centros educativos. Dicha información estaba relacionada con la detección de patrones o perfiles de alumnos con diferentes niveles de riesgo, asociados al bajo rendimiento y/o al abandono de los estudios.

La alternativa utilizada fue la *simulación neurocomputacional* del campo interdisciplinario de la Inteligencia Artificial, en atención a que su estructura y funcionamiento, como se expondrá ampliamente en el segundo capítulo, muestran características que la hacen idónea para realizar las tareas que implican identificar y reconocer patrones y secuencias de acciones, conjuntando así los aspectos de diagnóstico y prescripción psicológicos.

Para el presente trabajo se procuró ampliar y profundizar el horizonte de la investigación, buscando alcanzar dos metas importantes:

1) Incorporar al análisis psicométrico del reconocimiento de patrones, algunos desarrollos que actualmente gozan de prestigio internacional y que se basan en el modelo identificado de manera general como *IRT* (Item Theory Response), el cual ha recogido lo substancial de los trabajos de Rasch (1960/80) y ha tenido avances notables en autores como Embretson (1983, 1984, 1985, 1991), Andrich (1985, 1988), Lord (1980) y Wrigth y colaboradores (1977, 1979, 1982), entre otros.

2) Restar la artificialidad, que es casi inevitable, en las condiciones de exploración y desarrollo del entorno psicológico neurocomputacional, por ejemplo en las tareas experimentales, procurando acercarlas a la situación cotidiana en los ambientes escolares de la educación media superior y superior. En el presente estudio, el interés se centró en el reconocimiento de patrones de desempeño académico, sometiendo a prueba dos sistemas de procesamiento neurocomputacional, para contar con un contexto de análisis más poderoso.

El objetivo principal de este estudio es lograr la identificación de tres patrones diferenciales del desempeño académico (alto, medio, bajo) en muestras de sujetos egresados del nivel de secundaria, usando una doble vía: la psicométrica, enmarcada en la teoría *IRT* y la tecnológica, basada en el procesamiento neurocomputacional.

La identificación de patrones tiene como base los resultados de la aplicación de un instrumento similar a los que se emplean en nuestro medio, para evaluar conocimientos y habilidades que se asumen necesarios para ingresar al bachillerato, siguiendo la estructura del CENEVAL (1994).

La tesis propiamente dicha consiste en asumir, y someter a prueba, que el reconocimiento de patrones inherente al diagnóstico psicológico, en una amplia gama de variantes y modalidades, requiere primero de un análisis psicométrico cuidadoso para establecer los parámetros de interés y segundo, que una vez que han sido determinados, esta actividad puede hacerse con confiabilidad y velocidad considerables, haciendo uso de procedimientos automatizados *inteligentes*, que funcionan como expertos. Por esta razón, a estos procedimientos computacionales se les adscribe a un campo de especialización llamado apropiadamente Inteligencia Artificial.

En los primeros dos capítulos se presentarán, a manera de marcos de referencia, las perspectivas que ofrecen la aproximación *psicométrica* y la *neurocomputacional* a la identificación y el reconocimiento de patrones. En el tercero se describirá el problema de investigación abordado en este trabajo, para proceder después, en el cuarto capítulo, a señalar los aspectos metodológicos de este estudio, en relación a los sujetos, los materiales e

instrumentos, el diseño de la investigación y los procedimientos. En los capítulos finales se expondrán los resultados, discusión y conclusiones.

Más que pretender exponer en forma extensa y detallada los numerosos aspectos técnicos que se observan en los modelos *IRT* del dominio de la psicometría y en el campo híbrido de la Inteligencia Artificial (que ha tenido sus principales aplicaciones en la línea ingenieril), se optará por formar un marco conceptual que permita apreciar y contextualizar su utilización en nuestra área. En especial se hará una descripción sistemática de los procedimientos para favorecer la comprensión de las estrategias de desarrollo y prueba psicométrica y neurocomputacional.

Por último, cabe señalar que la *neurocomputación* es simplemente un método de análisis poco conocido en nuestro medio, lo que puede ocultar el valor generalizado de su utilidad. Consiste en un sistema muy poderoso para reconocer patrones, identificar y reconocer rasgos, características, tendencias, etc. que no está circunscrito a alguna disciplina en particular y su aplicación es común en materia de ciencias biomédicas, comunicaciones e ingeniería, entre otras.

En la presente investigación, esta tecnología innovadora permitió realizar exitosamente tareas propias del diagnóstico psicológico.

CAPITULO I

La perspectiva psicométrica en tareas de reconocimiento de patrones

a) Introducción

Thorndike y Lohman (1990) citan el trabajo de Cheng: *Psychological tests found in China during Han and Wei dynasties*, publicado en 1928, quien refiere que, mucho antes de la era cristiana, ya se usaban las pruebas de verdadero falso, completar dibujos y responder preguntas verbales. El desempeño era considerado indicación de inteligencia y educación.

Virtualmente, cada sociedad ha observado que existen diferencias de alguna clase entre sus miembros y se ha interesado por resaltarlas, aplicando algún calificativo que distinga una condición, rango o función. Para estos fines ha ideado pruebas de lealtad, valor, patriotismo, astucia, entre otras. Esto indica que el objeto de la evaluación y las tareas seleccionadas para "medir" diversos atributos asociados a los sujetos, dependen de los contextos culturales y de los valores vigentes.

Thorndike y Lohman (op. cit.) ofrecen en su obra un panorama ilustrativo de los esfuerzos realizados durante un siglo en la medición de habilidades en el campo de la psicología, particularmente en los Estados Unidos, aún cuando se reconocen los trabajos pioneros de Binet (Francia) y Spearman (Inglaterra). El análisis de su reseña sugiere que se ha tratado de un camino lento, árduo, lleno de retos y con frecuencia cuestionado, pero pródigo en aspectos interesantes, entre los que se encuentran:

1890 - Cattell llamó *pruebas mentales* a la investigación sobre fenómenos mentales.

1894 - Estudios de Gilbert sobre *sensación, tiempo de reacción y ejercicios de memoria*.

1895 - La APA crea una comisión especial encargada de las *pruebas mentales*.

- Trabajos de Binet y Henri sobre *la psicología individual*.

- Se funda *L'Année Psychologique*.

1896 - Ebbinghaus completó sus pruebas.

1899 - Las pruebas aplicadas a los niños de escuelas elementales fueron llamadas, en términos de la APA, *Pruebas Kirkpatrick*.

1901 - Los estudios de Wissler no encontraron correlación entre *tiempo de reacción y calificaciones escolares*.

- 1902 - Los documentos de Thorndike critican las pruebas en uso.
- 1903 - La tesis de doctorado de Damaye delineó los temas estándares para determinar el *retraso mental* de algunos individuos.
- 1904 - Thorndike publicó su libro *An Introduction to the Theory of Mental and Social Measurements*.
 - Spearman introdujo la teoría de dos factores (*g*).
- 1905 - Binet y Simon publicaron su primera escala.
- 1906 - Norsworthy recomienda elaborar *perfiles de habilidades*.
- 1908 - Binet y Simon publican una segunda escala.
- 1909 - Binet publica *Les idées modernes sur les enfants*.
- 1910 - Ayres publica *Laggards in Our Schools*.
- 1911 - Binet y Simon publican la revisión final de sus escalas.
- 1912 - Stern introduce el término *Intelligence Quotient*.
 - Spearman introduce la *ecuación tetracórica*.
- 1914 - Kong publica las *pruebas no verbales para inmigrantes*.
- 1915 - Yerkes introduce la *Point Scale*.
- 1916 - Yerkes y Wood introducen el término *Coefficient of Intelligence*.
 - Terman publica la Revisión de Stanford a la *escala de Binet-Simon*.
 - Thompson señala la correlación artificial que produce el *factor g*.
- 1918 - Pruebas de inteligencia *Army Alpha / Army Beta*.
- 1919 - Terman propone aplicar anualmente pruebas a grupos de alumnos.
- 1920 - Thomson introduce su *teoría de muestreo*.
- 1921 - Stoddard publica *The Revolt Against Civilization*.
 - Se desarrollan las *Pruebas Nacionales de Inteligencia*.
- 1922 - Lippmann difunde una serie de artículos criticando el uso de las pruebas.
- 1926 - Thorndike introduce la *teoría de cuatro dimensiones del intelecto*.
- 1931 - Thurstone introduce el *método del análisis factorial*.
- 1933 - Hotelling hace la generalización del método de Pearson de *análisis de componentes principales*.

- 1935 - Alexander en sus estudios encuentra *g*, algunos *factores de grupo* y dos factores de *inhabilidad*.
- 1937 - *Formas L y M* de la escala de Stanford-Binet (segunda edición).
- 1938 - Thurstone presenta un estudio de *Habilidades Mentales Primarias*.
- 1939 - Se presenta la escala *Wechsler-Bellevue*. Se introduce la *desviación IQ*.
- 1941 - Prueba de Clasificación General de la Armada (AGCT).
- 1943 - Cattell propone los factores *Gf* y *Gc*.
- 1947 - *Batería de Prueba de Aptitudes Generales (GATB)*.
- *Pruebas de Aptitudes Diferenciales*.
- 1949 - *Escala de Inteligencia Wechsler para Niños (WISC)*.
- 1953 - Pruebas de Flanagan *para la Clasificación de Aptitudes*.
- 1954 - La APA recomienda estándares para el desarrollo y uso de las pruebas.
- 1955 - La escala *Wechsler-Bellevue* es reemplazada por la *Escala de Inteligencia de Adultos (WAIS)*.
- 1959 - Modelo Guilford de *la estructura del intelecto*.
- 1960 - *Formas L-M* de la escala de Stanford-Binet (tercera edición).
- 1967 - Escala Wechsler de *inteligencia para preescolar y primaria (WPPSI)*.
- 1970 - Prueba de *Habilidades Cognoscitivas* (después conocida como el *CogAT*).
- 1972 - Tercera edición de la *escala de Stanford-Binet* con nuevas normas.
- 1973 - Primer estudio de Hunt sobre la *correlación del procesamiento de información con la inteligencia*.
- 1974 - Publicación de la revisión del WISC (*WISC-R*).
- 1977 - Sternberg presenta su *subteoría componencial de la inteligencia*.
- 1980 - Estudios de Jensen sobre *velocidad mental*.
- 1981 - Publicación de la revisión del WAIS (*WAIS - R*).
- 1982 - Publicación del *Handbook of Human Intelligence* de Sternberg.

1983 - Kaufman presenta su *batería de medición para niños (K - ABC)*.

1985 - Horn presenta su revisión de la teoría *Gf / Gc*.

- Sternberg presenta su teoría *triádica de la inteligencia*.

1986 - Cuarta edición de la *escala de Stanford-Binet*.

Reyes Lagunes (1992) realizó un estudio que muestra la panorámica de la evaluación educativa en México, con el propósito de señalar los roles asumidos en el trabajo del psicólogo: diagnóstico, propuesta y evaluación de programas, aplicaciones teóricas, en un periodo amplio que va de los años sesenta a los noventa, aunque centra su revisión en la última década.

Para la autora citada, la evaluación puede ser conceptualizada como *un proceso metodológico de investigación sistemática* que permite, entre otras cosas, realizar un diagnóstico, para conocer la naturaleza y extensión de un problema. Gracias a este proceso, la toma de decisiones en el campo educativo puede ser razonada y justificada, con independencia de las eventualidades teóricas o de otro tipo, que estén presentes en un momento dado.

Reyes Lagunes (op. cit), encuentra en su revisión cuatro grandes rubros que agrupan los desarrollos en evaluación educativa de acuerdo a su objetivo. El primero de ellos es el de los *diagnósticos*, que cubre un amplia gama de tópicos y niveles educativos.

En el presente trabajo, se coincide en que formular diagnósticos es en general una actividad que, en esencia, reproduce el proceso sistemático de la investigación en sus componentes principales, ya que realizarla implica crear una representación del problema, generar hipótesis que guíen la búsqueda de indicadores y someter a prueba las suposiciones. Constituye una actividad básica porque de ella depende, como ya fue citado, la posibilidad de tomar decisiones correctas.

Cuando en el marco educativo o instruccional se pretende evaluar algún factor (cognoscitivo, conductual, afectivo o social) que se estima que está asociado al desempeño escolar, se emplea un proceso metodológico que sigue el esquema trazado en relación al reconocimiento de patrones en la solución de problemas y su expresión más formalizada puede lograrse con el apoyo de técnicas psicométricas apropiadas.

No obstante que exista claridad en la tarea, hay numerosos problemas que es común encarar en su realización y conviene tenerlos presentes. Tal vez el más importante es que las escalas que se usan, por ejemplo para medir conocimientos, y habilidades, son inexactas, por caracer de unidades regulares; su significado y calidad dependen del conjunto específico de reactivos que se haya estandarizado y de los conocimientos y habilidades de los sujetos que estén representados en la muestra.

Wright y Stone (1979) preguntan: ¿cómo hacer mediciones objetivas y construir una ciencia del desarrollo intelectual, cuando se trabaja con reglas de hule? Estos autores ilustran su punto de vista de la manera siguiente: si se parte del número de reactivos que un sujeto acierta, puede obtenerse su posición, por ejemplo percentilar, en comparación con los demás sujetos evaluados con la misma prueba.

Pero ¿cómo deben interpretarse los resultados, más allá de los límites de ese conjunto particular de reactivos y de esos sujetos? Si cambian los sujetos se tendrá una nueva regla para medir; si cambian los reactivos de nuevo se tendrá otra regla para la medición. El problema real es que cada colección de reactivos mide una habilidad por su cuenta y es evidente que cada medida depende, para su significado, de su propia familia de examinandos.

Embretson (1985) señala que la medición psicológica y educacional han sido ampliamente criticadas en numerosos artículos y trabajos que conciernen a la validez de una prueba. Aunque los editores de las pruebas hacen referencia a correlaciones débiles con *criterios sociales relevantes*, las críticas se relacionan directamente con las habilidades, conocimientos y procesos involucrados en la solución de los reactivos de un examen.

La respuesta actual a algunos de estos problemas ha tomado como base los modelos de Rasch (1960/1980), un matemático danés, que se enfrentó a la tarea de estandarizar un grupo de pruebas de lectura y al problema de definir la dificultad de un reactivo, independientemente de la población meta y de la habilidad específica de un individuo, sin importar cuáles reactivos había resuelto.

Rasch no siguió la psicometría clásica y a cambio tomó la alternativa de aplicar el álgebra a un modelo probabilístico. El supuesto que observó en sus trabajos es ahora una posición clave en los modelos derivados de la Teoría de la Respuesta a los Reactivos (*IRT*): la probabilidad de que una persona resuelva correctamente un reactivo debe depender de un parámetro de *habilidad* pertinente sólo a la persona y de un parámetro de *dificultad* pertinente sólo al reactivo.

Rasch (op. cit.) desarrolló el *Modelo Logístico de Probabilidad para el Análisis de Reactivos*, razonando que podría lograrse la medición objetiva al aplicar una función llamada *logística* al análisis de los datos de las pruebas, tomando como centro la dificultad de los reactivos. El propósito práctico de sus procedimientos es la tecnología que sugiere para construir instrumentos de evaluación, cuya precisión dependerá sólo de la dificultad del reactivo, considerada óptima cuando se tiene 0.50 en la probabilidad de acertar.

Este planteamiento se ha beneficiado por una cantidad considerable de investigación, que aportó sustento teórico y soporte empírico. Andersen (1977) y Bandorff-Nielsen (1978), en trabajos independientes probaron que

lo único que puede ser estimado con consistencia y suficiencia a partir del análisis de las respuestas correctas e incorrectas disponibles, es la *dificultad* de un reactivo.

Esta tesis tomó la orientación de partir de la estimación de la dificultad de los ítems, agregando un segundo parámetro: estimar también el grado de habilidad de los sujetos, como ejes principales para alcanzar las metas de identificar y reconocer patrones de respuesta subyacentes al desempeño. En lo fundamental se apoyó en las aportaciones de dos autores destacados:

1. Embretson (1983,1984, 1985, 1991) de la Universidad de Kansas, quien conjuga en el diseño de pruebas de conocimientos y habilidades académicas dos componentes esenciales para el éxito en esta tarea: a) una teoría cognoscitiva robusta, como sustento para la medición de procesos complejos y b) la aplicación de modelos psicométricos apropiados a la complejidad de la tarea.

2. Wrigth y cols. (1977, 1979, 1982) de la Universidad de Chicago, quienes ofrecen un marco teórico-metodológico psicométrico que ha logrado impactar a la medición psicológica actual, a la vez que hacen accesibles los modelos logísticos para analizar reactivos, aplicando más de un parámetro, mediante sistemas computarizados que facilitan enormemente esta tarea.

Debe señalarse, también, la importancia de los trabajos de Andrich (1985, 1988), que explica con claridad los aspectos fundamentales de los modelos de medición Rasch; Thorndike (1989) que integró en una obra didáctica la amplitud y complejidad de la psicometría moderna, y de Lord (1980), que ha mostrado diversas aplicaciones de la teoría *IRT* a la atención de problemas de medición.

A continuación se presentan principios y componentes relevantes de los modelos que siguen la Teoría de Respuesta de Reactivos, abreviada *IRT*.

b) Principios y componentes de los modelos IRT.

Embretson (1985) señala que, en la aproximación tradicional a la medición psicológica, no ha habido suficiente fundamentación teórica, ni se ha tenido apoyo de los modelos psicométricos que podrían hacer explícitas las cualidades substanciales de los estímulos de prueba, o de las personas que responden las pruebas.

Campbell y Stanley (1970) hicieron algunas observaciones en el mismo sentido, en un material muy conocido en nuestro medio, Se trata de una revisión metodológica que particulariza el problema de la validez, justificando este interés al señalar que se trata de prevenir los efectos amenazantes de doce factores que atentan contra la inferencia válida.

Los autores mencionados citan varios trabajos de los años veinte, como el de McCall (1923), quien señaló que si bien se disponía de muchos libros y tratados acerca del "buen manejo" de los datos experimentales, se carecía, en cambio, de aquellos que orientaran sobre cómo obtener datos adecuados y correctos, a los cuales aplicar el procedimiento estadístico.

Puede advertirse que, desde ese tiempo, contribuía a la problemática en la investigación educacional el asumir que la práctica que no era fruto de la *experimentación* era ineficaz, conviviendo con la aversión a experimentar, por tratarse de una tarea que se sentía tediosa, equívoca, casi irreplicable y basada por lo común en conocimiento precientífico; en suma, frustrante. Acerca de esta situación, Campbell y Stanley afirman que lo más doloroso no es en realidad el proceso experimental en sí, ya que la verdadera fuente de la frustración es *partir de una teoría inadecuada*.

Para Embretson (1985) la solución al problema de la validez y de otros problemas relacionados, se encuentra en los avances más recientes en psicología y psicometría, construidos en el marco de la Teoría de la Respuesta de Reactivos. A estos desarrollos también se les identifica como *modelos de atributos latentes*, que ofrecen a la medición un conjunto de estrategias agrupadas bajo el rubro metodológico de *diseño de pruebas*. La suposición básica que subyace a esta aproximación, es que las cualidades a medir (los atributos latentes), referidas a un modelo teórico que sustente su inclusión, pueden elicitar cuando se aplican técnicas para diseñar y construir los estímulos de una prueba.

Se trata de modelos *proceso-producto* que son diseñados para evaluar las demandas cognoscitivas que plantean los reactivos de una prueba, ligando un modelo psicométrico a uno matemático. El propósito esencial es operacionalizar los constructos de una teoría, en particular la cognoscitiva de procesamiento de información, asumiendo que la respuesta de un reactivo requiere de información correcta, proveniente de los diversos componentes del procesamiento.

Esta característica procede del impacto de la psicología cognoscitiva sobre la evaluación, como ha sido documentado en abundantes trabajos de la literatura psicológica. Son modelos empleados internacionalmente en la construcción de bancos de reactivos y para la igualación de puntajes en las pruebas entre las diferentes versiones o subconjuntos de reactivos.

En el presente trabajo un objetivo general, como ya ha sido planteado, es ofrecer respuestas a la pregunta: *¿cómo construir pruebas apropiadas para evaluar conocimientos y habilidades en dominios especializados, que permitan identificar los patrones de respuesta que subyacen al desempeño académico, sin que los resultados dependan de quién mide, a quién se mide y con qué se mide?*

De manera resumida, puede decirse que el interés básico es disponer de un método que permita pasar de observaciones en las respuestas que se dan en una prueba, a mediciones precisas acerca de los conocimientos y habilidades de un estudiante. Lo que se requiere es un método de medición específico a la evaluación propuesta e independiente de la variación de otras características de los sujetos medidos, o de los instrumentos de medición empleados.

De acuerdo a Thorndike (1989), la información que se obtenga a partir de una prueba debe referirse, por lo menos, a dos componentes: las personas y los reactivos. En lo relativo a las personas, es claro que el interés principal es obtener evidencia confiable y precisa acerca de la magnitud de los conocimientos y habilidades para resolver las preguntas u operaciones, como las que están contenidas en el examen

Los trabajos de Embretson (1983, 1984, 1985, 1991) contienen varios planteamientos acerca de los modelos de la respuesta a los reactivos (IRT), que señalan problemas importantes en la medición de los conocimientos y habilidades involucrados en la evaluación del desempeño escolar.

1. Un conjunto de mediciones a partir de una prueba puede ser útil tanto para medir diferencias individuales como para representar *constructos* teóricos. Sin embargo, el diseño de una prueba requiere que la teoría sustantiva y los métodos psicométricos se apliquen conjuntamente al elaborar los reactivos. Dicho de otra forma, para comprender las respuestas que se dan a los reactivos, se necesita una teoría de las variables que están involucradas. Por lo tanto, la teoría psicológica (y su experimentación) es un aspecto fundamental del diseño.

2. Los métodos psicométricos que permiten encadenar propiedades sustantivas de los reactivos a las diferencias individuales, son igualmente críticos para el diseño de pruebas. Las variables que pueden influir en las respuestas, no contribuirán necesariamente a las diferencias individuales. Por lo tanto, se requiere un modelo psicométrico para clarificar cómo la manipulación de un aspecto específico de la tarea contenida en el reactivo, afectará los atributos que son medidos a través del desempeño.

3. Se requieren métodos psicométricos capaces de medir los efectos de variables independientes múltiples, porque es común que las teorías psicológicas postulen que más de una variable influye en el desempeño de una tarea. Para ser útil al diseño de una prueba, el impacto de cada variable debe ser medido por reactivos específicos.

4. El diseño de pruebas está muy relacionado con la validez de constructo, la cual depende tanto de su representación, como del espacio nomotético.

La *representación del constructo* concierne a los aspectos teóricos que explican las respuestas, al realizar una tarea en un reactivo. Para ello puede descomponerse un ítem en las tareas que lo integran. Al seleccionar reactivos acordes al impacto relativo de varios constructos, pueden diseñarse pruebas para representar aspectos específicos de las diferencias individuales.

Un constructo puede tener, o no tener, implicaciones importantes para medir diferencias individuales. Los sujetos pueden no variar en el constructo o éste puede estar altamente correlacionado con otros constructos en la población meta. Por lo tanto, también es muy importante estimar el *espacio nomotético del constructo*, determinando la frecuencia, patrón y magnitud de las correlaciones de los puntajes de las pruebas con las otras variables.

5. En general, la validez de constructo dependerá de la influencia relativa de las variables cognoscitivas que subyacen a las diferencias individuales en la solución de reactivos. Las pruebas con diferentes pesos en las variables cognoscitivas tendrán diferentes espacios nomotéticos.

6. Los desarrolladores de pruebas suelen ignorar las suposiciones teóricas relevantes a los reactivos y a sus métodos de investigación. En cambio, tienden a ser expertos en áreas de contenido específico. Las especificaciones para escribir reactivos son guías generales (estructura, formato y reglas acerca de contenidos específicos), por ejemplo, temarios, actualizaciones, etc. Otras veces la guía se reduce a la colocación de la opción correcta en el reactivo y a indicar lo que no se deben usar como opciones aceptables.

7. La especificación de las variables de dificultad y calidad de los reactivos tiene poca o ninguna base teórica. Sin embargo, la experiencia de los elaboradores de reactivos, en conjunción con los análisis psicométricos de las respuestas, suelen generar pruebas con un buen nivel de validez predictiva. Este es el eslabón de las pruebas con la teoría psicológica y ha sido una solución para la validez de constructo.

8. Los problemas que se encuentran en la evaluación institucional, por ejemplo para la admisión de aspirantes, han llevado a tomar decisiones que obedecen a presiones públicas o administrativas. Por ejemplo, proponer exámenes "únicos", a pesar de que se carezca de un modelo de evaluación teórico unificado, de manera que los instrumentos son híbridos contruidos con reactivos que difieren en aspectos importantes.

Un componente, también crítico, suele ser la ausencia de equivalencia de los reactivos en diferentes versiones del examen. La equivalencia de los reactivos puede basarse en el procedimiento llamado *calibración de ítems*. Las características de los estímulos en los reactivos, los constructos que operacionalizan y sus valores psicométricos son la base para la igualación.

9. El contenido de las pruebas puede afectar al menos en dos sentidos, los aspectos de las diferencias individuales que son medidas:

Primero: un conjunto pequeño de ítems, aunque sea una selección óptima para proporcionar información eficiente acerca del nivel de habilidad de la persona, puede no ser representativa de las diversas facetas de la tarea ejecutada, cuando se le compara con el banco total.

Segundo: la mezcla de reactivos en un subconjunto puede cambiar las demandas de la tarea. El contexto de un subconjunto específico puede ser muy poco representativo de los reactivos del banco como un todo.

10. La *habilidad escolar*, según es medida en las pruebas, puede ser conceptualizada como una variable interviniente, que resume las relaciones entre variables cognoscitivas que subyacen al procesamiento y las medidas externas. Los pesos de las diversas variables cognitivas en el puntaje de la prueba determinan la naturaleza de la habilidad. Por ello, *la habilidad* está eslabonada al diseño de pruebas.

11. Las características de los estímulos en los ítems equivalen a variables que controlan los componentes que están involucrados en el desempeño. Estas se combinan con una o más estrategias aplicadas por los sujetos para resolver el reactivo. La habilidad para responder depende en parte de la ejecución en varias estrategias. Por lo tanto, los parámetros de los reactivos pueden usarse para diseñar pruebas que midan aspectos específicos de las diferencias individuales.

12. Los modelos psicométricos multicomponentiales que miden atributos latentes, proporcionan estimados de las demandas cognoscitivas en cada reactivo y especifican las relaciones de estas demandas con las habilidades cognoscitivas, que se reflejan en la solución de los reactivos. Este contexto tiene la ventaja incidental de proporcionar parámetros para la persona y para el reactivo, en cada variable cognoscitiva postulada en el modelo.

13. Los componentes principales de los *Modelos Multicomponentiales de Atributos Latentes* (MLTM) han mostrado ser útiles para probar hipótesis acerca de los procesos cognoscitivos involucrados en la realización de las tareas que plantean los reactivos de prueba. Por ejemplo, el modelamiento de componentes cognoscitivos en reactivos de pruebas de inteligencia, como reportan Sternberg y McNamara (1985) y Pellegrino, Mumaw y Shute (1985), puede ser realizado en el contexto de un modelo psicométrico. Este contexto tiene la ventaja incidental de proporcionar parámetros para la persona y para el reactivo en cada variable cognoscitiva postulada en el modelo.

Embretson (1985) hace una revisión amplia de trabajos que ofrecen evidencia de la aplicación de los principios psicométricos mencionados,

abordando una diversidad temática que muestra el alcance de los supuestos teóricos. Entre ellos se encuentran los que enfocan la comprensión de las pruebas de aptitud al explicar las variables cognoscitivas que determinan la ejecución. Los métodos para identificar el impacto de las variables teóricas en una tarea (por ejemplo, de composición) se consideran cruciales para ligar "aptitud" y "teoría cognoscitiva". Para ilustrar cómo se descompone una tarea en tipos específicos de reactivos, se hace referencia a los trabajos de Sternberg y McNamara, Pellegrino et. al., y Butterfield et. al., publicados en 1985.

Los tipos de reactivos seleccionados para estos trabajos incluyen vocabulario, construcciones espaciales y completamiento de series. En ellos están representados tres modalidades principales de habilidad: verbal, espacial y razonamiento inductivo. Además ofrecen sumarios de la literatura cognoscitiva que muestra hallazgos relevantes. También ilustran el trabajo de los autores al diseño de variables para el tipo de reactivo que puede ser implementado en la construcción de pruebas. Dado que se muestra cómo descomponer una tarea particular para ilustrar los principios básicos, su comprensión ayudará a que se generalice a otros tipos de reactivos y a otros métodos de descomposición de tareas.

Sternberg y McNamara (op. cit.) descomponen la representación y el procesamiento cognoscitivo de reactivos de vocabulario, asumiendo que es un índice de la inteligencia verbal, que ha recibido poca atención. Estos autores introducen innovaciones, al operacionalizar dentro de sus modelos matemáticos diversas teorías de la representación cognitiva del significado de las palabras. Muestran cómo varían los reactivos en la representación de los atributos característicos y en la definición de atributos, lo que a su vez influye en la dificultad de los diversos eventos de procesamiento que están involucrados en la solución de reactivos. También examinan estrategias para el procesamiento de vocabulario, que son definidas por las secuencias y tipos de comparaciones de la palabra meta con las alternativas.

Embretson menciona que el trabajo de Sternberg y McNamara tiene dos implicaciones mayores para el diseño de pruebas. Primero, las variables de la teoría de estos autores acerca del análisis de reactivos de vocabulario pueden ser usados para controlar la dificultad de la representación y procesamiento de cada reactivo. Segundo, la validez de constructo de reactivos de vocabulario se explica claramente al comprender las variables de representación y procesamiento involucradas en la respuesta a los ítems. Esta información puede ser utilizada para evaluar el tipo de ítem como un todo, con respecto a las metas propuestas para la prueba.

Entre los puntos de interés del trabajo de Pellegrino, Mumaw y Shute (1985) destacan sus métodos para descomponer tareas. Estos resultados incluyen un análisis paralelo de los tiempos de respuesta y de los errores de

respuesta. Si se combinan ambos aspectos, el análisis se hará un tanto complejo, por lo que se recomienda analizar los componentes por separado.

El *modelamiento del error* parece ser más relevante para las metas de la evaluación, dado que los índices psicométricos más empleados (como el puntaje total), está relacionado inversamente con el número de errores. Los resultados contestan la pregunta de cuáles procesos contribuyen a los errores en la ejecución y, por tanto, cuáles procesos son importantes en las diferencias individuales en habilidad. Si embargo, los datos del tiempo de respuesta son necesarios para fundamentar una variable teórica como un evento de tiempo real en el procesamiento.

Butterfield, Nielsen, Tangen y Richardson (1985) desarrollaron un modelo que predice la dificultad en las tareas de completamiento de series. Esto es importante porque se mide el razonamiento inductivo en tareas relativamente libres de contenido, que requieren razonamiento básico. Y son útiles para predecir con precisión el éxito, permitiendo anticipar la dificultad del reactivo antes de los datos empíricos.

Snow y Swanson (1985) examinaron el diseño de pruebas desde un nivel diferente al de los precedentes, ya que se interesaron más por las variables molares. Esto es, revisan las manipulaciones que influyen en la ejecución sobre un bloque de reactivos relacionados y no se interesan por las diferencias entre reactivos individuales. Postulan que lo que se refleja en un puntaje sobre un conjunto de reactivos depende substancialmente de las condiciones que influyen el conjunto total, tales como formato, instrucciones, práctica o estrategia de entrenamiento.

Estos autores sugieren que la combinación de la interacción aptitud-tratamiento (ATI) y de métodos que analizan el procesamiento cognoscitivo de información, es una condición necesaria para comprender mejor las bases cognoscitivas de los puntajes. Analizan ejemplos de estudios en los cuales las condiciones globales, tales como el permitir la toma de notas durante el examen, o el proporcionar reactivos más concretos, influyen los procesos cognoscitivos que se reflejan en la puntuación. Esta manipulación puede introducirse de manera sistemática, de manera que las diferencias de ejecución entre dos condiciones pueda ser puntuada para representar los procesos que se manipulan.

En otro trabajo, Embretson (1991), refina conceptos relacionados con los componentes cognoscitivos representados en un constructo y hace referencia al diseño y la aplicación de modelos psicométricos más complejos y poderosos para contender con el reto que las diferencias individuales plantean a la evaluación educacional.

Embretson asume que la habilidad responsable de una ejecución, llamada "*habilidad latente efectiva de un individuo*", es la suma de varias sub habilidades; una de ellas se tenía antes de observar un cambio. Cada una

de las habilidades que resultan del aprendizaje representan un incremento efectivo entre dos mediciones sucesivas.

Este modelo muestra características metodológicas sustantivas para la investigación psicológica. Un rasgo es la claridad: cuando el modelo se ajusta apropiadamente a los datos, se asegura la comparabilidad del ractivo y la prueba. Esto permite construir instrumentos comparables que contengan diseños de mediciones variadas para usarse en otras ocasiones.

Esto es importante, porque lo deseable en muchas clases de estudios es que se pueda disponer de mediciones diferentes, pero comparables, del mismo atributo. También es útil para hacer mediciones repetidas, con base en formas alternativas de examen. Similarmente, cuando los individuos son medidos repetidamente a lo largo de sus vidas (estudios longitudinales), es necesario disponer de diferentes conjuntos de pruebas. La propuesta de Embretson ayuda a que el mismo atributo pueda ser medido con precisión y comparar los cambios.

Medir habilidades a partir de los progresos en el aprendizaje ha tenido un atractivo especial, desde el comienzo de la medición moderna de las habilidades. En las principales conferencias sobre *inteligencia*, muchos participantes han enfatizado que la capacidad para el aprendizaje es su mejor indicador. Sin embargo, como se ha señalado desde hace ya muchos años, la mayoría de las pruebas en uso no miden la capacidad para aprender, sino lo que ha sido *aprendido*.

El interés en la medición del aprendizaje continúa en la investigación contemporánea y, en las teorías sobre la *inteligencia*, se puntualiza que la visión de las pruebas *estáticas* de la habilidad para el aprendizaje se basa en la suposición (fallida) de que los procesos requeridos para el desempeño en una prueba están también directamente envueltos en el aprendizaje.

Al margen de este interés permanente, los intentos que se han hecho, para medir las diferencias individuales en la habilidad para aprender, han sido ampliamente inefectivos. La revisión de la literatura sugiere que tanto los problemas sustantivos, como los psicométricos, conducen a fallar en la medición de las habilidades para aprender.

La aproximación actual a la medición de habilidades para aprender difiere sustancialmente en varios aspectos de los estudios realizados en los años 30 (por ejemplo: Woodrow, 1938), que asumían que las diferencias individuales en los cambios debidos a la práctica, eran *específicos a la tarea de aprendizaje* y que por ello no eran predictores válidos en contextos educativos. En nuestros días, las habilidades escolares se miden con tareas complejas, muy cercanas a la predicción del éxito. Las condiciones para medir se diseñan con el propósito de evaluar los cambios en el desempeño, que se desprenden más de procesos cognoscitivos específicos, que del desconocido impacto de la práctica.

Por ejemplo, Sternberg (1977) ha desarrollado modelos teóricos para comprender las fuentes de la complejidad cognoscitiva en los reactivos y, aún más, ha propuesto métodos que permiten manipular la dificultad de las fuentes subyacentes. Las habilidades para aprender están referidas a las fuentes de la complejidad cognoscitiva, que son influenciadas por las claves para resolver los reactivos y por el entrenamiento previo a la condición de medición.

Embretson (1991) concluye que las principales dificultades en la medición clásica se refieren a: (1) que tanto el nivel medio del cambio como su significado (medido por las correlaciones) dependen de la distribución de los puntajes iniciales y (2) que esta distribución es específica a la población. La persistencia de estos problemas ha originado su reconceptualización en la Teoría de Respuesta de Reactivos, ensayando nuevas soluciones.

De la revisión antes presentada pueden destacarse varios puntos de interés para los objetivos de este trabajo, entre los que se encuentran:

1. Los modelos multicomponenciales de atributos latentes permiten medir las demandas cognoscitivas de los reactivos, al ligar un modelo psicométrico a uno matemático que operacionalice los constructos de una teoría, en particular de procesamiento de información. El énfasis de la evaluación está puesto en la respuesta a cada reactivo y esto con el fin de precisar diferencias individuales, como objetivo primario.

2. Los modelos *IRT* asumen que para resolver un reactivo se requiere de información correcta proveniente de los diversos componentes del procesamiento. Esto es más obvio cuando se aplican en particular a la elaboración y el análisis de reactivos complejos, por ejemplo para medir habilidades intelectuales, como son: analogías verbales, silogismos, completamiento de series, emisión de juicios y razonamiento matemático.

3. Para estimar las demandas cognitivas en un conjunto de reactivos, debe haber una teoría explícita de los procesos subyacentes, que ofrezca un buen ajuste a los datos. Las variables cognoscitivas tienen niveles distintos y pueden afectar diferencialmente la puntuación total de una prueba. Muchas fuentes de diferencias individuales al procesar información, no se exploran con la profundidad debida. Por ejemplo, Sternberg (1977) contempla varias de ellas en su teoría *componencial*. Sin embargo, el enfoque principal en la mayoría de los trabajos que evalúan conocimientos y habilidades está puesto sólo en la puntuación total en la prueba.

4. Algunos reactivos pueden ser resueltos aplicando más de una estrategia, que pueden no haber sido consideradas en la representación del constructo. Una estrategia involucra una *combinación de componentes* que pueden ser ejecutados para resolver el reactivo. El total de reactivos puede

ser resuelto con distintas estrategias, dependiendo de las decisiones que tomen los sujetos para ejecutarlas, lo que sesga la evaluación.

5. Las combinaciones de los pesos de los estímulos y las estrategias determinan la naturaleza de la habilidad medida por la puntuación en la prueba y, por lo tanto, también se debe determinar su espacio nomotético.

c) Aspectos Procedimentales de los modelos IRT (*)

Wright y colaboradores (1979, 1982) han trabajado en elaborar manuales para hacer medición Rasch y en la construcción de programas computarizados para apoyar su desarrollo. En secciones introductorias a estos materiales hacen algunos señalamientos importantes acerca de los problemas involucrados en la medición psicológica y educativa, en particular elaboran el punto relacionado con la generación de medidas no arbitrarias.

Para estos autores, la medición comienza con la idea de una variable o una línea, a lo largo de la cual puedan ubicarse los objetos, con el propósito de hacer marcas en esta línea en unidades iguales, de manera que permita comparar la distancia entre dos puntos.

Los objetos en la medición educativa se refieren a personas y los números que se derivan de ellas son mediciones. Por lo tanto, la medición de una persona es un estimado de su posición en la línea que representa la variable.

Los instrumentos para la observación son pruebas, cuestionarios y reactivos y los números que se derivan de ellos son calibraciones, para significar su papel instrumental en el proceso de medición. La calibración de un reactivo es un estimado de su posición en la línea de la variable a lo largo de la cual las personas son medidas.

Se mide a las personas y se calibra a los reactivos sobre la variable que definen conjuntamente. Por lo tanto, la construcción de una variable requiere de una relación sistemática y reproducible, entre personas y reactivos.

En este modelo básico, el planteamiento central es que la respuesta a un reactivo está dada por un nivel de probabilidad de acertar o de fallar, que solo depende de la habilidad de la persona y de la dificultad del reactivo.

(*) El desarrollo de esta sección se benefició considerablemente de las orientaciones surgidas en discusiones metodológicas y de comunicaciones personales para definir estrategias de evaluación, mantenidas en diversas ocasiones con el Dr. José Manuel Álvarez Manilla, precursor del análisis de Rasch en México.

El énfasis del análisis es determinar cómo responde cada persona los reactivos y a partir de este análisis transformar los datos crudos a unidades *logic*, que permiten pasar a una escala normalizada, a partir de la cual pueden establecerse los patrones de respuesta para cada persona y cada reactivo. A la medida en que hay ajuste o desajuste del nivel de habilidad a la dificultad del reactivo y de éste a la habilidad de las personas se le concede un valor clave y muy significativo, que de hecho soporta toda la estrategia metodológica.

Puede considerarse que en general se busca descubrir valores de ajuste como indicadores del funcionamiento de cada reactivo, lo que lleva a tomar decisiones sobre si deben conservarse o eliminarse de un instrumento dado y tomar las respuestas de los sujetos como indicadores para tomar decisiones con efectos educacionales.

Algunos desajustes provienen de que un sujeto acierta al responder un reactivo, cuando su patrón haría esperar que fallara; y de que falle en un reactivo cuando su patrón daría probabilidades altas de que acertara.

El inicio de estos procedimientos es tomar como datos crudos las respuestas a los reactivos para generar matrices de datos que implican un reordenamiento y hacer transformaciones para determinar matemáticamente el mejor valor de ajuste.

Estos planteamientos acerca del diseño de pruebas, la construcción de escalas y sobre la medición educacional, encierran propuestas centradas en dos aspectos: a) la construcción de variables para medir atributos latentes y b) cómo medir las variables para hacer comparaciones objetivas, obtener predictores y deducir suposiciones bien establecidas para apoyar la investigación psicológica.

A continuación se señalarán algunos principios y técnicas útiles para realizar la medición, que proponen Wright y Stone (1979).

1. Cuando se dice que un sujeto está en el percentil 90, por ejemplo en habilidad matemática, necesitamos saber en cuál grupo y de qué prueba se habla, para que tenga sentido. Cuando se afirma que ese sujeto tiene una estatura de 1.89 Mts, se preguntan Wright y Stone (1979) ¿tenemos que pensar en cómo se obtuvo la medida? El instrumento de medición puede variar en color, peso, composición y tamaño, pero su escala es independiente de todos estos aspectos. Se asume que esta medición es objetiva, tan es así que la persona medirá lo mismo cuando se le mida con otra regla y esto será independiente de quien haga la medición. En cambio, dependiendo del grupo con el que se compare, tendrá un percentil de habilidad diferente.

A la medida que posee esta propiedad se le llama *objetiva* y para que la adquiera son necesarias dos condiciones: a) la calibración de

instrumentos de medición debe ser independiente de los objetos empleados para la calibración, y b) la medición de objetos debe ser independiente de los instrumentos usados para la medición.

Se enfatiza particularmente que la construcción de modelos de medición con esa clase de objetividad debe aceptar que la calibración de la dificultad reactivo-prueba no puede depender de las personas específicas medidas para hacer la calibración y que la medición de una habilidad no puede depender de los reactivos específicos empleados para la calibración.

2. Cuando se compara un reactivo con otro para calibrar una prueba, no debería importar cuáles respuestas se usarán para la comparación. El método para calibrar exámenes debe dar los mismos resultados, con independencia de las personas a quienes se desea evaluar. Esta forma asegura que se puedan construir instrumentos uniformes e independientes de los sujetos.

3. Cuando se mide la habilidad de una persona mediante un conjunto de reactivos, no debe preocupar cuál selección de ítems debe responder. La meta es comparar personas y llegar a establecer medidas equivalentes estadísticamente, siendo indistinto los reactivos empleados y el momento en que fueron medidas con exámenes completamente diferentes.

4. Tradicionalmente, la calibración de una prueba se hace observando cuántas personas aciertan al responder un reactivo. Esta visión de *dificultad* corresponde a la proporción de respuestas correctas. La calidad de un ítem se juzga, entonces, por la correlación entre las respuestas y la puntuación total en la prueba. La habilidad de una persona se indica por su ubicación en un percentil, en la misma muestra. Sin embargo, un problema frecuente proviene de las suposiciones concernientes a la adecuación de la muestra de personas, que se tomó para estandarizar una prueba.

Es posible intentar una nueva aproximación, en la que no se requiera hacer suposiciones acerca de la distribución de la habilidad de las personas y que se base en un modelo de lo que ocurrirá al responder un reactivo.

5. El modelo de lo que ocurrirá al responder un reactivo, señalan Wrigth y Stone (1977), dice simplemente que el resultado dependerá por completo de *la diferencia entre la habilidad de la persona y la dificultad del reactivo*. Mientras más hábil sea la persona, mejorarán sus oportunidades de responder el reactivo correctamente. Mientras más fácil sea el reactivo, más personas responderán acertadamente.

Este modelo permite tomar en cuenta cualquier habilidad que interese de las personas en la calibración de la muestra y estimar la dificultad de la prueba desde las particularidades de estas habilidades.

Para obtener una medida es necesario que se tenga una variable en la cual se pueda ubicar dicha medición. Una variable comienza con la idea

general de lo que se desea medir y toma substancia cuando se empiezan a diseñar y elaborar preguntas de examen dirigidas a elicitarse señales o indicadores de la variable en las respuestas. Si la variable (conocimientos, habilidades, aptitudes) se visualizara como una línea, entonces la medida podría ser representada como un punto en esa línea, según se muestra en la siguiente figura

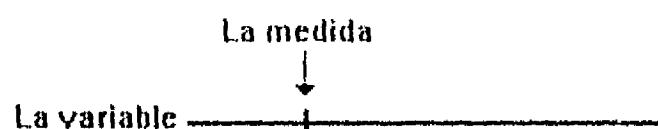


Figura 1.1. Representación de una variable y una medición

Cuando una persona es sometida a una prueba, el propósito es estimar su localización en la línea, implicada por la prueba. Para que esto sea posible, se debe contar con una prueba que defina la línea. También debe tenerse un procedimiento para convertir las respuestas de una persona en la prueba, a una localización en la línea que representa la variable de interés. Por lo tanto, las preguntas de examen pueden usarse para definir las líneas y las respuestas a esas preguntas para ubicar en dichas líneas a las personas.

Para construir una prueba que defina una variable (conocimientos y/o habilidades), las preguntas o reactivos que formen la prueba deben permitir el establecer una *línea de inspección*, como se muestra en la siguiente figura:

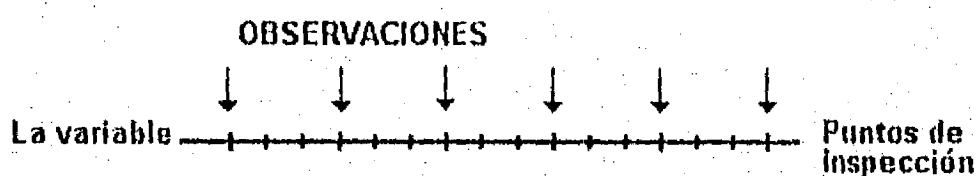


Figura 1.2. Puntos de inspección de una variable.

Esta línea debe marcar una dirección (\Rightarrow) que permita determinar los valores que corresponden a las observaciones. Por ejemplo, para examinar con facilidad la magnitud del dominio del conocimiento o habilidad que se mide. El significado de esta línea está dada por las preguntas de examen que la definen.

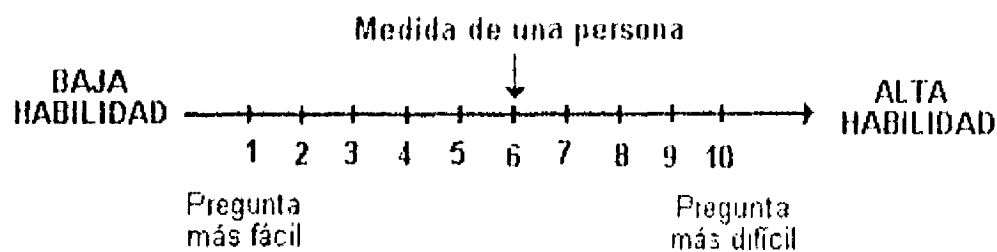


Figura 1.3. Dirección de las observaciones que definen una variable.

Las marcas en la línea representan a cada uno de los reactivos y, en este ejemplo el ordenamiento obedece a grados de dificultad crecientes en los ítems. *Las marcas son la calibración* de las preguntas a lo largo de la variable y estas preguntas calibradas constituyen la *definición operacional* de lo que la variable mide. Los reactivos aportan la definición operacional de la variable porque son indicadores precisos y objetivos de los conocimientos o habilidades que se desean medir.

Las preguntas más *difíciles* permitirán identificar a las personas más *hábiles* y definirán el extremo más alto, a la derecha de la línea. Asimismo, las preguntas más *fáciles* identificarán los niveles de habilidad más *bajos*, en la porción izquierda.

La experiencia del evaluador en la generación de preguntas de examen y en la construcción de instrumentos de medición es algo esencial. También lo es el aprender a calibrar los reactivos para tener instrumentos apropiados.

A estos planteamientos habría que agregar que toda medición de conocimientos y habilidades tiene ligado un error de medición y que el tamaño de éste deberá estimarse con precisión. Ambos valores, el de la medida y el de error, estarán presentes en los resultados.

Asimismo, partiendo de la necesidad de disponer de información acerca de las personas que responden la prueba y de las preguntas que integran los instrumentos, puede asumirse que en la aplicación se conjuntarán ajustes y desajustes en ambos sentidos: personas por arriba o por debajo de la dificultad media de un examen y reactivos por encima o por debajo del nivel medio de habilidad de las personas.

Wright y Masters (1982) señalan que hay varios métodos que fueron desarrollados para ser usados en cinco diferentes formatos de respuesta, y que en conjunto son el corazón del modelo psicométrico contenido en los modelos IRT, formando una familia que han hecho posible la calibración de reactivos: 1. *Dichotomus Model*, 2. *Poisson Counts Model*, 3. *Binomial Trials Model*, 4. *Rating Scale Model* y 5. *Partial Credit Model*.

Todos ellos tienen formas algebraicas semejantes y pueden ser pensados como extensiones del modelo básico de Rasch, quien partió de su aplicación al análisis de datos *dicotómicos*. En este modelo básico, el planteamiento central es que la respuesta a un reactivo está dada por un nivel de probabilidad de acertar o de fallar, que solo depende de la habilidad de la persona y de la dificultad del reactivo.

d) Componentes de una estrategia metodológica para la evaluación integral del desempeño académico (*)

López y Castañeda (1992) diseñaron un *Modelo Multicomponencial y Multidimensional para la Evaluación Integral del Aprendizaje Universitario* y buscaron consolidarlo incorporando hallazgos derivados de la investigación propia y de la que fue generada de la revisión bibliográfica de corrientes vanguardistas en psicología y psicometría cognoscitivas, fundamentando teórica y empíricamente su estructura y funcionamiento.

Este modelo ha sido adoptado por la UNAM para aplicarlo a la selección de aspirantes a ingresar al bachillerato, al nivel técnico de enfermería y a las cuatro áreas disciplinarias de la licenciatura.

Se trata de un modelo integral, que aportó elementos esenciales de la realización de este trabajo, en particular para:

a) enmarcar el estudio en las líneas teóricas que exploran la cognición humana, fuente de los constructos de *habilidad* y *conocimiento* manejados en este trabajo;

b) generar una taxonomía con capacidad probada para calibrar la dificultad y complejidad de las preguntas de examen y medir con precisión los indicadores de los constructos incorporados, en su interacción particular con los dominios de contenido específico (asignaturas);

c) orientar la construcción de bancos de reactivos por docentes expertos en las materias evaluadas, aplicando estrategias para diferenciar el desempeño académico, muestreando las demandas de las tareas implícitas en los reactivos, y

d) categorizar los reactivos atendiendo a su asociación con los niveles de habilidad de los sujetos, permitiendo la identificación de patrones de respuesta diferenciales y el reconocimiento de los perfiles de desempeño.

(*) Este texto es una versión sintetizada del *Modelo Psicológico de Asesoría Instruccional para el Desarrollo Cognoscitivo*, elaborado por López y Castañeda (1993), para el Proyecto de la UNAM: Procedimiento de Evaluación para el Ingreso al Bachillerato,

En el desarrollo de esta propuesta fueron muy enriquecedores los trabajos de Anderson (1983) y Anderson y Bower (1973), Gagné (1985) y Gagné (1987). A continuación se ofrece una descripción somera de las características relevantes de este modelo que está actualmente en la fase de refinamiento.

Dimensiones del Modelo de Evaluación del Desempeño Académico.

1. *Cognoscitiva*: identifica la influencia de las estructuras de los diversos dominios de conocimiento, precisa las demandas cognoscitivas de las tareas a realizar en la evaluación y define los tipos, amplitud y complejidad de los conocimientos y habilidades a evaluar.

2. *Edumétrica*: precisa los fines pedagógicos de la evaluación, es decir: qué y para qué se evalúa.

3. *Psicométrica*: aporta el marco de referencia metodológico para englobar los aspectos técnicos de validez, confiabilidad y equiparabilidad en los reactivos y en los instrumentos de evaluación.

En el presente modelo, la evaluación del desempeño en tareas académicas es vista como un proceso *continuo, adaptativo, sistemático e integrador*, cuya artificialidad se reduce al incorporar elementos que satisfacen criterios de validez y confiabilidad.

Siguiendo a Thorndike (1989), el modelo conjunta tres componentes:

1. *Los atributos de las personas que contestan las pruebas*
2. *Las pruebas que intentan medir esos atributos*
3. *Los reactivos o preguntas de examen particulares que componen las pruebas.*

Por una parte, se trata de determinar la forma en que los atributos se relacionan entre sí, midiendo la capacidad o el dominio que se tiene en cuanto a los conocimientos y las habilidades de interés. Por la otra, se trata de dar cuenta de las relaciones que existen entre las propiedades de los reactivos componentes y la prueba, así como de las relaciones entre la calificación de la prueba y la propiedad del atributo subyacente.

Por su naturaleza, el modelo descrito puede ser situado en la tipología de procedimientos que se diseñan específicamente para evaluar el desempeño, con base en la medición de conocimientos y habilidades que se exhiben al responder un examen, como índice de lo que ha sido aprendido.

La estructura del modelo contempla los siguientes aspectos: 1) sensibilidad a las características de la población meta, 2) precisión al identificar y evaluar los conocimientos y habilidades académicas que se han asumido como básicos (por ejemplo: CENEVAL, 1994) para ingresar a

bachillerato y **3)** objetividad al fijar los niveles de dificultad apropiados para evaluar el desempeño.

Desde otro ángulo, el modelo pretende ser sensible también a la estructura y particularidades de cada dominio de contenido específico y a la complejidad de las actividades académico cognoscitivas, que están implícitas o explícitas en los programas de estudio.

La dificultad de los reactivos es vista como una *función interactiva* de los siguientes factores: **1)** los dominios de conocimiento definidos por las materias o asignaturas cursadas, que se muestrean para ser evaluados en el examen, **2)** el nivel de conocimientos previos del estudiante, **3)** la complejidad cognoscitiva de los reactivos, expresada en las habilidades requeridas para realizar las tareas señaladas en la prueba y **4)** los contextos de recuperación de lo aprendido.

En síntesis, el modelo atiende a diferentes variables relevantes, con el objetivo de alcanzar una evaluación integrativa, de alta confiabilidad, que favorezca el análisis cuanti y cualitativo de los resultados. Por todas estas propiedades y características se juzgó como idóneo para establecer perfiles de desempeño.

A continuación se identificarán los componentes del modelo y más adelante se describirá lo que enfatizan en la evaluación.

I. Contenidos básicos. Este rubro incluye tres aspectos:

1. Dominios del conocimiento. Se analizan cuatro niveles:

- A. Asignaturas (biología, español, física, etc.).
- B. Unidades (por ejemplo: evolución, corrientes literarias),
- C. Temas (por ejemplo: teorías de conjuntos, citoquímica, cinemática) y
- D. Subtemas (por ejemplo: conceptos de magnitud escalar y magnitud vectorial, estructura química de las enzimas, modernismo).

2. Tipos de conocimiento. Se distinguen tres variantes.

- A. Declarativo
- B. Procedimental
- C. Condicional

3. Niveles de conocimiento:

- A. Factual
- B. Conceptual
- C. Procedural

II. Habilidades escolares.

1. Generales:

- A. Identificar
- B. Inferir
- C. Resolver

2. Específicas. Relacionadas con:

- A. Aptitudes verbales.
- B. Razonamiento abstracto.
- C. Razonamiento matemático.

III. Taxonomía Cognoscitiva

El modelo incorpora taxonómicamente las siguientes dimensiones:

1. *procesos cognoscitivos*
2. *habilidades cognoscitivas generales y específicas*
3. *contenidos*
4. *niveles de conocimiento.*

1. Procesos Cognoscitivos

Aportan el marco general para contextualizar el constructo que interesa evaluar, centrado en diversos atributos latentes que se expresan en procesos que al formar parte de las funciones intelectuales más generales, constituyen la base de la vida académica.

Gracias a estas funciones, es posible procesar la información que proviene tanto del exterior como de nuestro medio interno y permiten que se realicen operaciones de aprendizaje complejas, como adquisición, transformación, recuperación y aplicación del conocimiento. Estos procesos generales se invocan cuando se emplean expresiones tales como *pensar, recordar, comprender, etc.*

La validez de *constructo* identifica en el modelo el espacio cognitivo contemplado en la evaluación, como base para definir las pruebas y determinar los niveles de complejidad de la ejecución, en términos de habilidades académicas, ya sea generales o específicas y de tipos y niveles de conocimiento, que toman formas concretas en operaciones intelectuales que son propias de actividades comunes en contextos escolares.

2. Habilidades académicas.

Las habilidades académicas o, como las llama de Sánchez (1992) las operaciones intelectuales, derivan de procesos o funciones más generales y enfatizan el manejo y la aplicación de información en actividades instruccionales, como son: *identificar, reconocer, ejecutar, comparar, indicar, ordenar, relacionar, clasificar, resolver, desarrollar, elaborar, organizar, estructurar, explicar, etc.* Es decir, se refieren a respuestas u operaciones que los reactivos demandan de quien contesta un examen y se expresan como instrucciones o indicaciones.

Para los propósitos del modelo se seleccionaron tres habilidades prototípicas amplias, inclusivas y generalizables: identificar, inferir y resolver.

Identificar es una forma elemental de recuperar la información adquirida, para poder reconocer algo previamente aprendido: hechos, conceptos, datos, fechas, definiciones, patrones, procedimientos, que se ajusta mucho al tipo de conocimiento declarativo (Anderson, 1983) y que en gran parte se apoya en procesos de la memoria.

Inferir es una categoría cognoscitiva amplia que se refiere al manejo de la información para establecer relaciones, por ejemplo entre hechos conceptos, principios y procedimientos. Por su naturaleza, involucra procesos deductivos, inductivos y analógicos. Entre las actividades académicas que incluye están el agrupar, definir, ordenar, secuenciar, separar, organizar, encontrar equivalencias, ya sea semánticas, funcionales o formales. Requiere conocimiento amplio del contenido de aprendizaje, por ejemplo de sus características y atributos, así como del análisis de los aspectos que se van a agrupar, diferenciar o integrar. Con esta habilidad es posible establecer causas, consecuencias, efectos y conclusiones derivadas de los planteamientos, aunque explícitamente no estén expresados. Por ejemplo, cuando se afirma: "*Si A es igual a B y B es mayor que C, se concluye que...*", mediante la Inferencia se llena el hueco de conocimiento.

Cuando se parte de una regla o principio aplicable a un hecho particular (deducción), o cuando este se toma para enunciar el principio a que corresponde (inducción), la habilidad básica es inferencial y se aplica al enunciar predicciones y encontrar relaciones, por ejemplo temporales, causales, jerárquicas, entre otras. Cuando las relaciones se establecen por similitud en los aspectos esenciales que se comparan, las respuestas se obtienen por analogía. La expresión: "*es como...*"; muestra el formato más típico de las analogías.

Resolver implica hacer uso del conocimiento adquirido y elaborarlo o procesarlo mediante el razonamiento, por ejemplo para analizar un

problema, aplicar un procedimiento y derivar un resultado. O bien, para seleccionar de entre varias posibilidades una estrategia o plan que permita llegar a resultados correctos, por ejemplo, reanalizar el planteamiento de una situación problema, decidir si es apropiado el empleo de una ecuación, examinar qué datos faltan o por qué falló el procedimiento, etc. En gran parte esta habilidad demanda que se organice o reorganice información, analizar relaciones entre hechos y conceptos y efectuar correctamente operaciones apropiadas. Este rubro conjunta los tipos de conocimiento seleccionados en este estudio, por su alta generalidad: *factual*, *conceptual*, *procedural* y sin duda habría que agregar además al *condicional* (Anderson, 1983), que son la base de la aplicación del conocimiento y el proceder algorítmicamente.

Debe señalarse que todas estas formas de habilidad son sensibles tanto al contenido específico (física, historia, etc.) como a los tipos de conocimiento que se incluyen en las tareas de evaluación y que de estas interacciones proviene el nivel de complejidad que puede tener un reactivo dado, además de los efectos que se derivan de otras variables, como los niveles de conocimiento previo.

En complemento a lo expresado en el párrafo anterior, cabe también hacer referencia a habilidades académicas *específicas*, que surgen de la *interacción* que se establece con la estructura y complejidad de distintos dominios de contenido, expresados como materias, asignaturas, áreas disciplinarias, campos temáticos, permitiendo así enfocar la evaluación a contextos instruccionales particulares. Por esa razón son relevantes, ya que contribuyen de manera más nítida a establecer el perfil académico del desempeño, partiendo del análisis de aciertos y errores. Sobre esta base es posible, incluso, prescribir acciones remediales con valor formativo.

Como ejemplos de habilidades escolares específicas pueden citarse:

- a) las que se refieren a *aptitudes verbales* (comprensión de lectura de varios temas y estructuras de texto; uso del lenguaje en expresiones formales atendiendo al manejo lexical, sintáctico y semántico, etc.);
- b) las que se relacionan con *aptitudes para el razonamiento abstracto* (ubicación espacial y temporal de acontecimientos y lugares relevantes; vincular factores y agentes con sus causas, consecuencias, efectos, implicaciones; tratar con hechos y fenómenos, distinguiendo en ellos sus propiedades, magnitudes y transformaciones); y
- c) las asociadas a *aptitudes para el razonamiento numérico* (efectuar operaciones de cálculo aritmético y matemático que varían en dificultad y complejidad; seleccionar un método cuantitativo que sea apropiado para solucionar un problema).

3. Contenido que se evalúa

Mientras que las habilidades académicas identifican algunas formas concretas que pueden tomar los procesos cognoscitivos, *el contenido* es la información sobre la cual operan. En el modelo desarrollado, el contenido se refiere a los diversos dominios específicos (español, física, matemáticas), denominados *asignaturas* y a sus componentes de menor inclusividad, como las *unidades, temas y subtemas*.

4. Tipos y niveles de conocimiento.

En conjunto constituyen indicadores útiles para poder identificar de manera precisa la información que está representada en los contenidos que se evalúan en los reactivos, partiendo de la conceptualización que señala Anderson (1983) de tres tipos generales.

A. Declarativo: engloba al conocimiento acerca del mundo, ya sea como conceptos, definiciones, descripciones, enunciados, declaraciones, acerca de qué son, cómo son, para qué sirven, cómo funcionan los objetos que nos rodean, y que se expresan mediante formulaciones verbales habladas o escritas.

B. Procedimental: es el mundo del saber hacer, el manejo de operaciones para que las cosas funcionen. Mientras que el conocimiento declarativo permite conceptualizar, por ejemplo, qué es un ácido y cómo se forma, el conocimiento procedimental o *procedimental* consiste en realizar todas las operaciones necesarias para formar un ácido. Otro ejemplo sería entender qué es *elegir un número al cuadrado* (conocimiento declarativo) y ser capaz de realizar las operaciones para hacerlo (procedimental). Conjuguar un verbo, resolver un problema, despejar una ecuación, son todos ejemplos de conocimiento procedimental.

C. Condicional: es el conocimiento de las condiciones necesarias para que las cosas funcionen o sean ciertas. Por ejemplo, qué cambios ocurrirán si se modifica un paso de un procedimiento, qué determina un resultado, a qué se debe que un resultado no sea correcto. En cierto modo, expresa la comprensión de por qué son así las cosas que nos rodean y de las condiciones que subyacen a sus transformaciones.

Con este marco de referencia en mente, en el modelo se juzgó conveniente incluir al conocimiento *condicional* en el marco más amplio del *procedimental* y de subdividir el de tipo *declarativo* en *factual* y *conceptual*, para identificar categorías de análisis que ayudaran a distinguir varios niveles en el manejo del contenido, apropiadas al material instruccional. La taxonomía incorporó tres agrupaciones altamente inclusivas: hechos, conceptos y procedimientos. Esto se tradujo en una organización con tres niveles de complejidad y dificultad del contenido.

Dados estos componentes, fue posible construir reactivos que combinan de varias formas la complejidad del contenido y los procesos y habilidades que inducen o imponen, a partir de las tareas y las instrucciones generativas. Así puede medirse el manejo de hechos, o de conceptos, principios y procedimientos, en contextos de prueba que demanden su identificación, inferir sus efectos o su aplicación. O por el contrario, se puede construir una evaluación que considere tan solo identificar un hecho en tareas de igualación a la muestra. Una y otra evaluación difieren sustancialmente.

Al incorporar al modelo de evaluación la taxonomía descrita, se creó un marco de trabajo útil para determinar el nivel de habilidad involucrado en la respuesta de un reactivo, calibrando teórica y empíricamente los niveles dificultad bajo, medio o alto, para identificar patrones diferenciales en la respuesta al examen.

Estos elementos dieron fundamento a la suposición central mantenida en el modelo: las *habilidades académicas* no deben ser evaluadas en abstracto, sino más bien como grados o niveles de la eficacia con que se realizan operaciones específicas (recordar, categorizar), expresadas en acciones concretas (ordenar, resolver, clasificar), al trabajar con contenidos específicos (español, química), para medir si se realizan en forma apropiada o incorrecta, como evidencia de su dominio.

CAPITULO II

La perspectiva neurocomputacional en tareas de reconocimiento de patrones

a) Introducción

La *Inteligencia Artificial* (IA) es un desarrollo interdisciplinario, surgido de las ciencias cognitivas (matemáticas, lógica, filosofía, lingüística, psicología), que en general sigue dos corrientes de pensamiento: a) reproducirse inteligencia como se pueda, con los conocimientos actuales que tenemos sobre las máquinas y b) reproducirse inteligencia artificial, tan cerca de lo que uno piensa que es la cognición humana (Negrete, 1990).

La neurocomputación es una derivación tecnológica de la ciencia cognitiva, que comparte características del dominio psicológico y de la IA y toma su nombre en analogía con la forma en que los sistemas biológicos transmiten la información, por ejemplo, en las estructuras del sistema nervioso central, que se valen de redes para la propagación de impulsos.

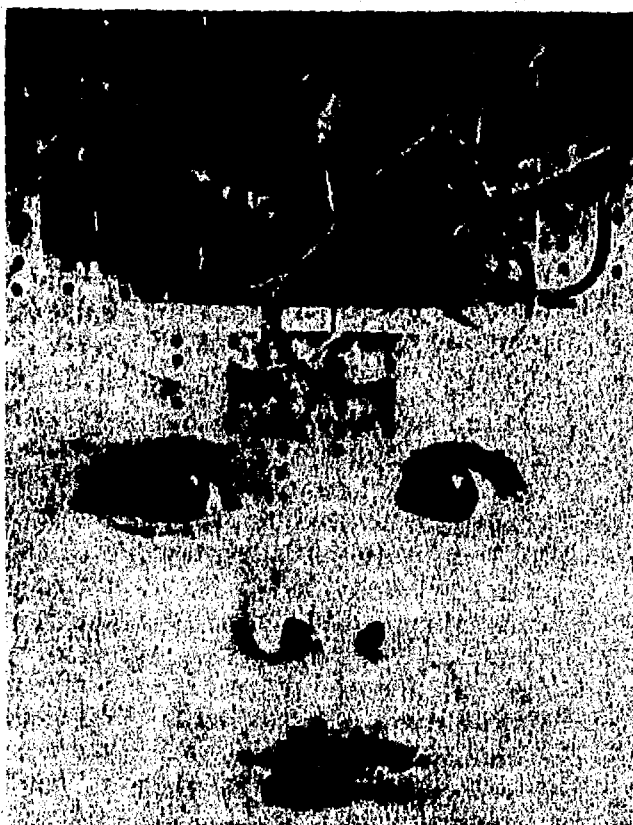
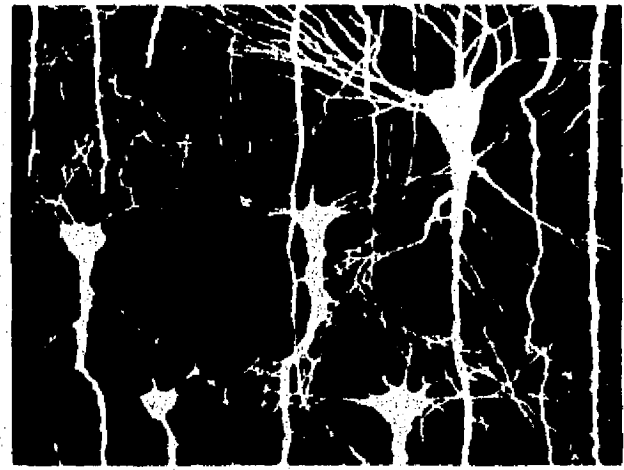
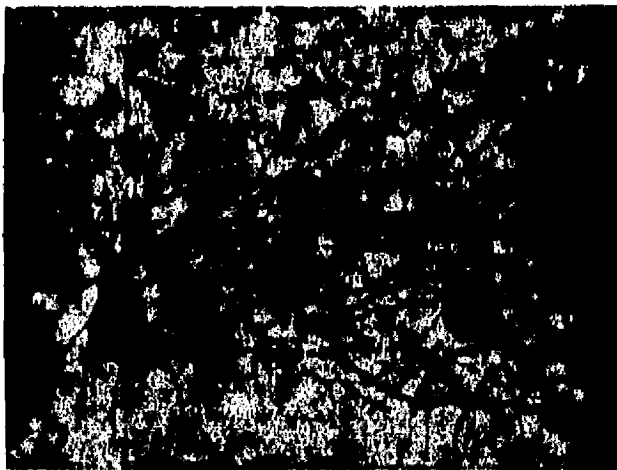


Figura II.1. El campo de la *Inteligencia Artificial* responde al reto de generar inteligencia, aunque sea artificial.

Feldman (1982) usó el término *conexionismo* para referirse a una clase de modelos que realizan sus cálculos a través de las conexiones entre unidades simples. Este paradigma también se conoce con el nombre de neurocomputación y de redes neurales. De estos modelos son en particular relevantes los de procesamiento distribuido en paralelo (McClelland, Rumelhart y el PDP Research Group, 1987); su interés principal es el aprendizaje y se asume que este representa la actividad natural del sistema.

En el campo de la neurocomputación los mecanismos básicos que ocurren en el cerebro y que se relacionan con pensar, recordar, aprender y solucionar problemas son concebidos como procesos de transmisión de información.

La transmisión de información en los sistemas biológicos, como se observa en las siguientes ilustraciones, es un proceso que opera mediante el intercambio de señales entre elementos o unidades, que forman redes muy densas. Se llama neuronas a estas unidades fundamentales. En los sistemas artificiales, los equivalentes funcionales de estas unidades se denominan *neurodas* y se asume que forman *redes neurodales*.



Figuras II. 2 y II.3. *Neuronas y redes neuronales.*

Una neurona actúa como un microprocesador relativamente simple. Es decir, como un mecanismo diseñado para recibir y combinar las señales de muchas otras neuronas. Las señales pueden variar en intensidad; esto es, pueden ser muy intensas o muy débiles. Si las señales combinadas que entran al cuerpo neuronal alcanzan un cierto valor, llamado umbral, entonces se produce una activación o disparo que genera una señal de salida (Espinoza, 1990; Carrillo, 1990).

La señal que dispara una neurona es recibida por las dendritas de otra neurona y si ocurre una sinpasis, ésta podrá retransmitir la señal a través de su axón a otra u otras neuronas con las que esté interconectada, dando lugar así a un proceso de propagación en una red. El resultado de esta actividad puede ser una respuesta sensitiva o motora.

En la simulación neurocomputacional, la analogía no se hace realmente con la estructura biológica, sino con el comportamiento de las neuronas, que reciben de otras la información, propagándola a su vez a través de redes complejas y extensas. Excepto por este hecho, de suyo muy significativo, no hay mayor relación entre ambos sistemas. Sus aplicaciones obedecen a los intereses de quien estudia y no por un campo específico. En todo caso, el funcionamiento del cerebro se parecería más a redes de computadoras muy potentes, trabajando todas ellas en paralelo.

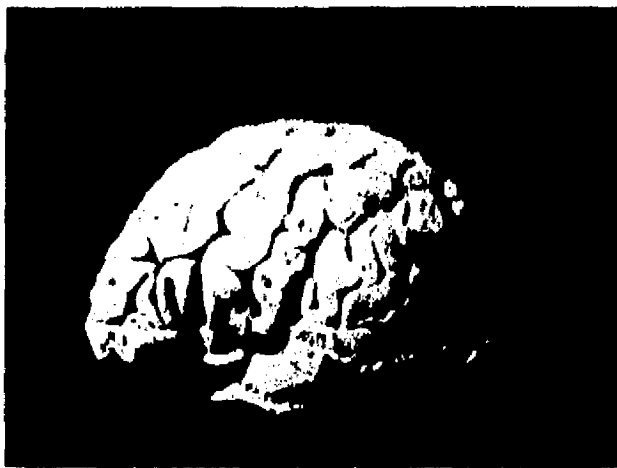


Figura II.4 *¿el cerebro es como una computadora?*

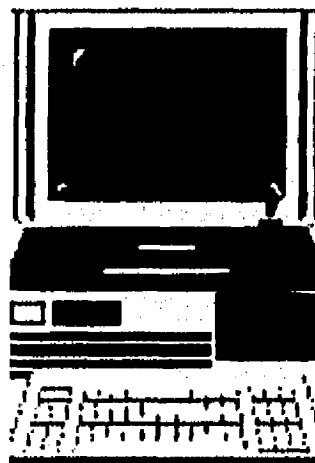


Figura II.5 *¿una computadora es como el cerebro?*

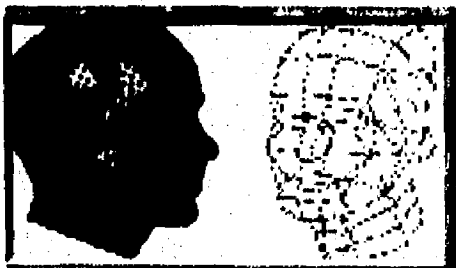


Figura II. 6. *¿Pueden ser simuladas las funciones cerebrales?*

En las estructuras artificiales estas funciones son simuladas mediante sistemas computacionales que operan de manera similar a como lo hace una red de comunicaciones. Su función primordial es organizar, propagar y

regular el flujo de información entre los componentes de la red, como en el sistema nervioso en relación a las neuronas y otras estructuras emisoras y receptoras de información. En ambos casos se trata de procesos regulados en que intervienen componentes de distinta naturaleza e importancia para el objetivo a lograr.

En síntesis, una red neurodal es un sistema computacional integrado por elementos sencillos y muy interconectados que procesan información gracias a su estado dinámico de respuesta a estímulos externos. Las unidades fundamentales, como se ilustra a continuación, son análogas a las neuronas y, como en los sistemas biológicos, sus interconexiones forman redes (Carrillo, 1990).

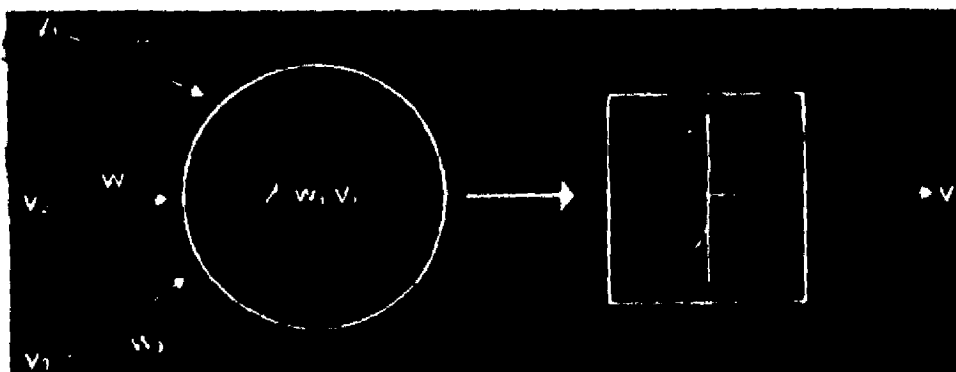


Figura II. 7. Una neuroda tiene una estructura muy similar a la de una neurona.

En la figura se indica que hay varias rutas (V_1, V_2, V_n) de acceso de la información (actuando como dendritas) y que cada señal que recibe de las otras neurodas con las que está interconectada tiene un peso o intensidad (W_1, W_2, W_n). Esto se traduce en niveles de actividad interna en el cuerpo neurodal, en este caso en respuesta al peso W_1 de la señal V_1 .

Cada unidad de procesamiento tiene capacidad para generar información propia y para difundir la que recibe de otros componentes, como un computador en miniatura. Pero para que entre en operación se requiere alcanzar el valor de un umbral. A estos valores se les llama pesos. La propagación se realiza gracias a una función de transferencia basada en: los pesos de las neurodas, el valor de los datos de entrada y el de los datos de salida que se pretenden igualar como meta o modelo.

Cuando estas señales se reciben en el interior de la neuroda, se activa un proceso que combina y suma los pesos para generar una respuesta, como se ilustra a continuación.

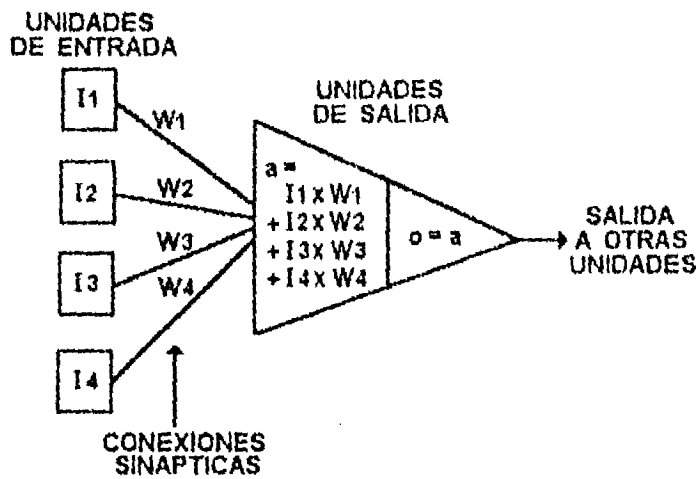


Figura II.8. Procesamiento en el interior de una neuroda.

Ahora bien, una red neurodal requiere que un cierto número de elementos de procesamiento estén interconectados, lo que por lo común integra grupos de neurodas llamadas *capas*. Una red típica consta de una secuencia de capas: entrada, intermedia y salida, como puede observarse a continuación.

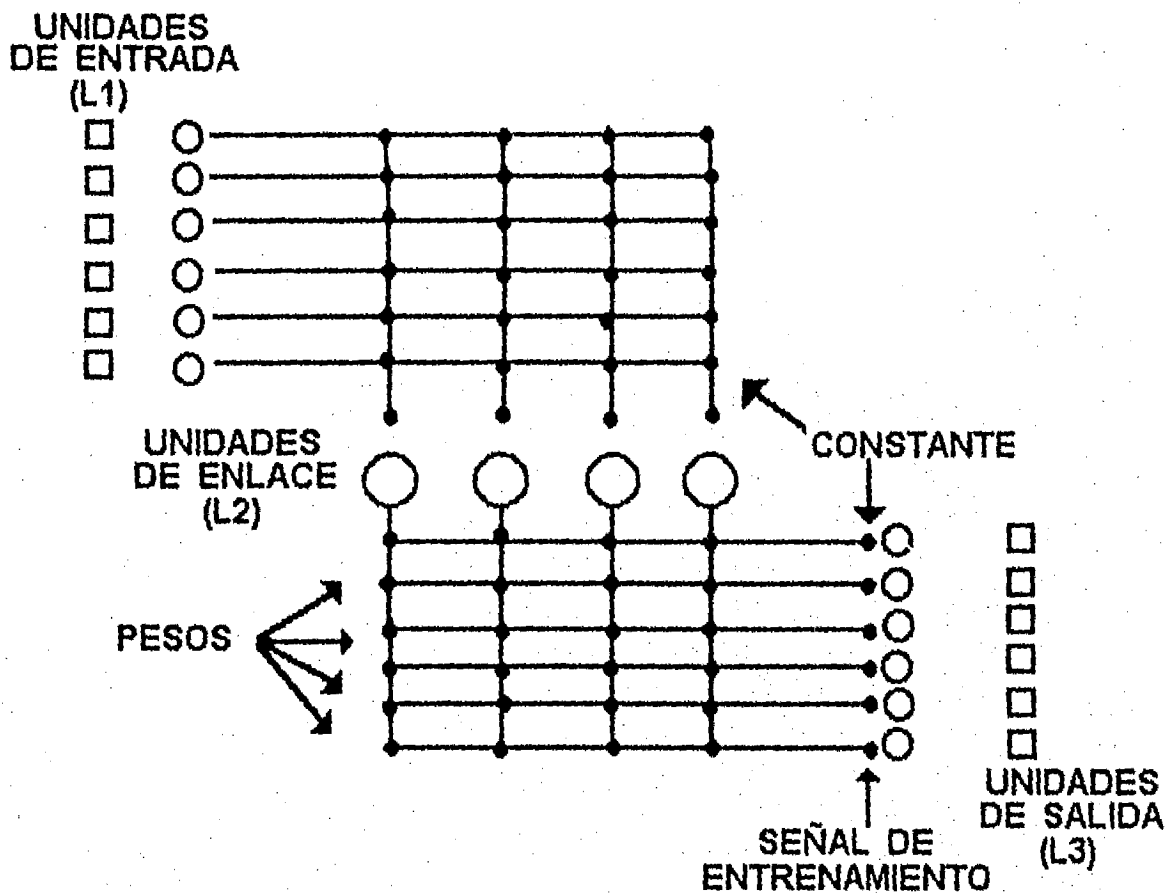
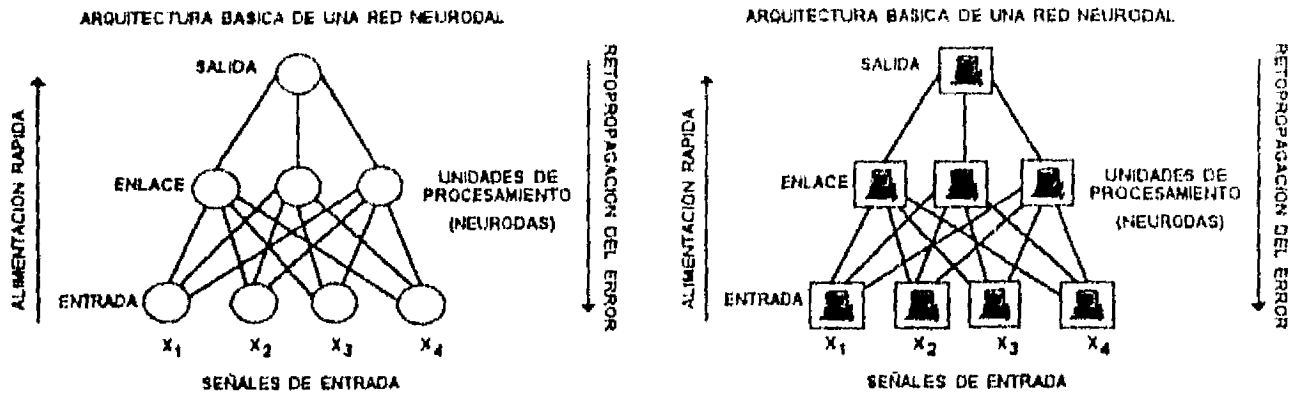


Figura II. 9. Estructura de una red neurodal en capas de procesamiento.

Dependiendo de cómo estén diseñadas las redes (arquitectura), sus neurodas pueden estar conectadas total o parcialmente entre sí, o incluso los enlaces entre capas sucesivas pueden obedecer al azar.

La siguiente representación muestra la estructura de una red simple con tres capas de procesamiento.



Figuras II.10 y II.11. *Arquitectura de una red neurodal. Las neurodas se integran en capas de procesamiento y cada unidad actúa como un microprocesador.*

En estas figuras se hace una analogía funcional con un conjunto de unidades, que en lo esencial se comportan como las computadoras. Se muestran tres capas de procesamiento. La primera de ellas es de *entrada* y recibe el flujo de las señales externas a la red, provenientes de otros sistemas. Algunas redes pueden recibir los valores de las señales que genera su capa de salida, dando lugar a un ciclo de realimentación. La segunda capa es de *enlace e interconexión* entre la entrada y la salida y es la responsable de proceder a la importantísima tarea de organizar y regular la información. La tercera capa es de *salida* y muestra los valores que genera la red.

La comparación continua de los valores de salida con los de entrada (patrón de entrenamiento) da lugar a ajustes en la capa intermedia, hasta que se reduce el error (tamaño de la diferencia entre entrada y salida), con base en un margen de tolerancia especificado (por lo general, entre 0.05 y 0.10).

También puede observarse que las neurodas de entrada y de salida no están conectadas directamente entre sí. Y que, en este ejemplo, las neurodas de una misma capa tampoco lo están. Sin embargo, gracias al sistema de interconexiones, los valores de entrada y enlace son compartidos a la o las neurodas de salida. El procesamiento es global y cada *lectura* involucra a todos los componentes de la red.

La información sigue dos flujos en su propagación: la que es necesaria para el entrenamiento de la red va hacia adelante: entrada-enlace-salida. En cambio, la de error va hacia atrás: salida-enlace-entrada. Lo que determina el valor de error son los pesos de los vectores de entrada y salida y su diferencia es lo que genera los ajustes. Como señalan

Rosenberg y Blalock (1988) el proceso de la propagación, ya sea hacia adelante o hacia atrás puede ser descrita en cuatro pasos básicos.

A. Propagación hacia adelante (entrenamiento):

1. Distribuir los valores de activación de las unidades a sus respectivas salidas.
2. Multiplicar las activaciones por los valores de los pesos.
3. Sumar estos valores de los pesos con los de la siguiente capa de unidades.
4. Aplicar la función logística a este valor.

B. Propagación hacia atrás (corrección del error):

1. Distribuir los valores de error de las unidades a sus respectivos pesos de entrada.
2. Multiplicar el error por los valores de los pesos.
3. Sumar esos valores de los pesos con los de la capa o unidades previas.
4. Evaluar la primera derivada de la función logística.

Cada unidad de procesamiento tiene capacidad para generar información propia y para difundir la que recibe de otros componentes, como un computador en miniatura. Pero para que entre en operación se requiere alcanzar el valor de un umbral. A estos valores se les llama pesos. La propagación se realiza gracias a una función de transferencia basada en: los pesos de las neuronas, el valor de los datos de entrada y el de los datos de salida que se pretenden igualar como meta o modelo.

La suma de los pesos combinados de las señales de entrada a la neurona puede ser modificada o ajustada por una función matemática llamada *función transferente*. Esta puede operar como un umbral, que sólo deje pasar información si la suma de los pesos alcanza un cierto valor. O bien, puede operar también de manera que permita el flujo de la información en forma continua. En una u otra variante, lo que genera es una señal de salida.

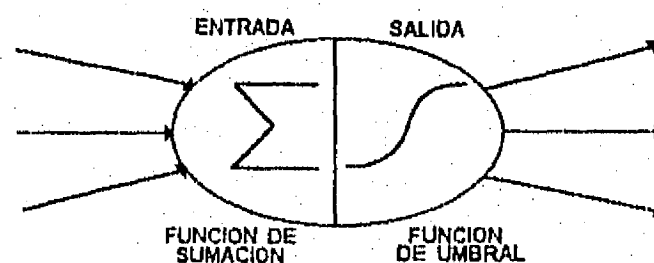


Figura II.12. Funciones de sumación y de salida de una neurona.

Dado que cada conexión tiene un cierto peso, las señales de entrada a una neuroda son modificadas por estos valores antes de ser combinados y sumados. Por lo tanto, hay una función de *sumación*, que toma en cuenta dichos pesos de entrada.

En esencia, una red neurocomputacional procesa información mediante modelos matemáticos que en su concepción elemental parecen un tanto simples: operar con valores entre 0 y 1 para decidir si se propaga o no la información de un componente a otro, con la capacidad de hacer ajustes en su funcionamiento, como se ilustra en la siguiente figura de una red con tres capas.

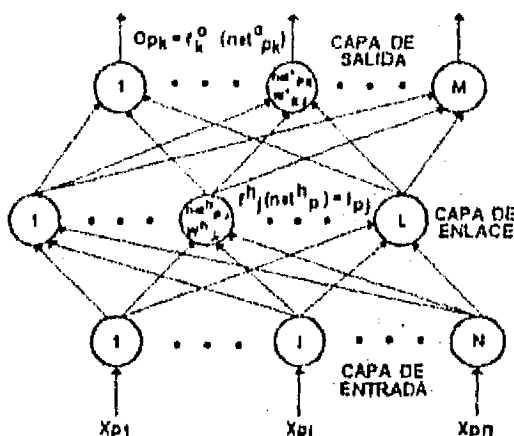


Figura II.13. Ajuste de los valores de salida mediante la sumación de los pesos

Una diferencia importante entre las redes neurodales y las biológicas es que las primeras son diseñadas para que cumplan funciones que van más allá de la transmisión de información, ya que pueden ser *entrenadas* para que *aprendan* ciertos tipos de relación y que *recuerden* los valores o parámetros aprendidos, de manera que al recuperar la información lo hagan con niveles de error cero o muy bajos.

Esta descripción de una red *neurodal computacional*, surgida del campo de la Inteligencia Artificial alrededor de los años 40, con capacidades para *aprender* y *recordar*, se asemeja a la que en el campo psicológico se podría hacer de una persona que aprende mecánicamente o por repetición.



Figura II.14. Un procesador neurocomputacional no requiere comprender lo que aprende.

En neurocomputación, el aprendizaje es concebido asociativamente como una adaptación o modificación de *pesos de conexiones*, en respuesta a los estímulos presentados como entradas. El proceso por el cual una red es inducida a aprender es visto como *entrenamiento*. Los estímulos (datos) que se presentan constituyen respuestas deseadas o esperadas a dichas entradas de información.

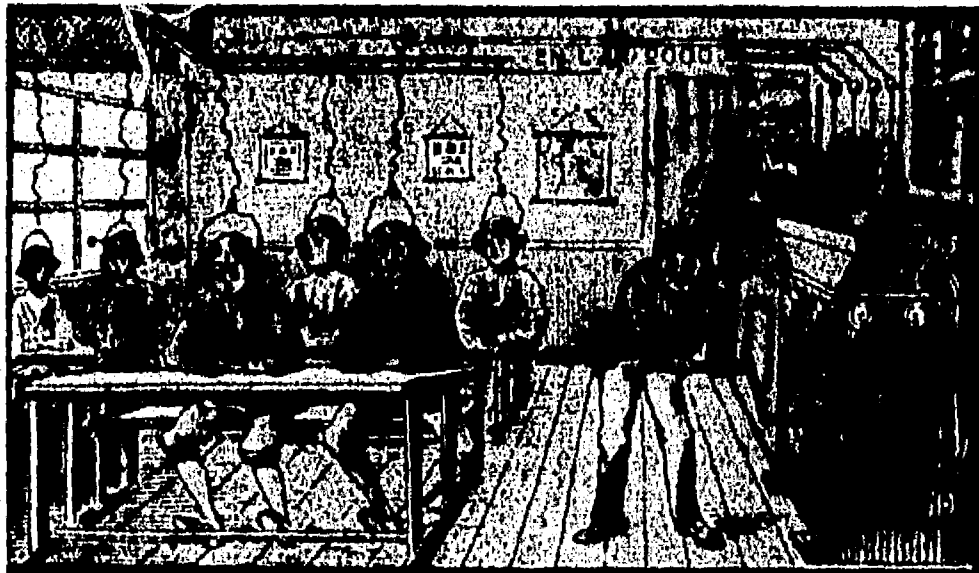


Figura II.15. Los estímulos de entrada son los patrones que se deben reproducir.

Indistintamente de qué clase de aprendizaje se trate, un rasgo común a cualquier red es que hay *reglas* que especifican cómo se adaptarán los pesos en respuesta a un ejemplo o modelo. Esto puede requerir que se presenten a la red muchos ejemplos varios cientos o miles de veces. Los parámetros que gobiernan una regla de aprendizaje pueden cambiar en el tiempo conforme la red progresa en el entrenamiento. El control a largo plazo que ejercen estos parámetros se conoce como *programa de aprendizaje*. La mayoría de los sistemas conexionistas usan alguna variedad de la regla Delta o de la regla Delta modificada, gracias a la cual ajustan

sus pesos hasta que logran aprender, por ejemplo, que las unidades comprometidas reproduzcan un patrón, como se ilustra a continuación.

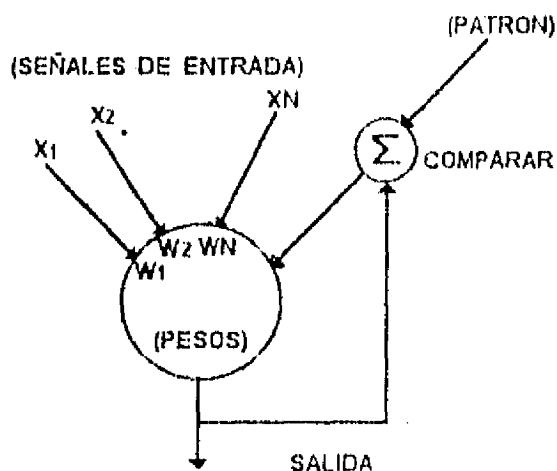


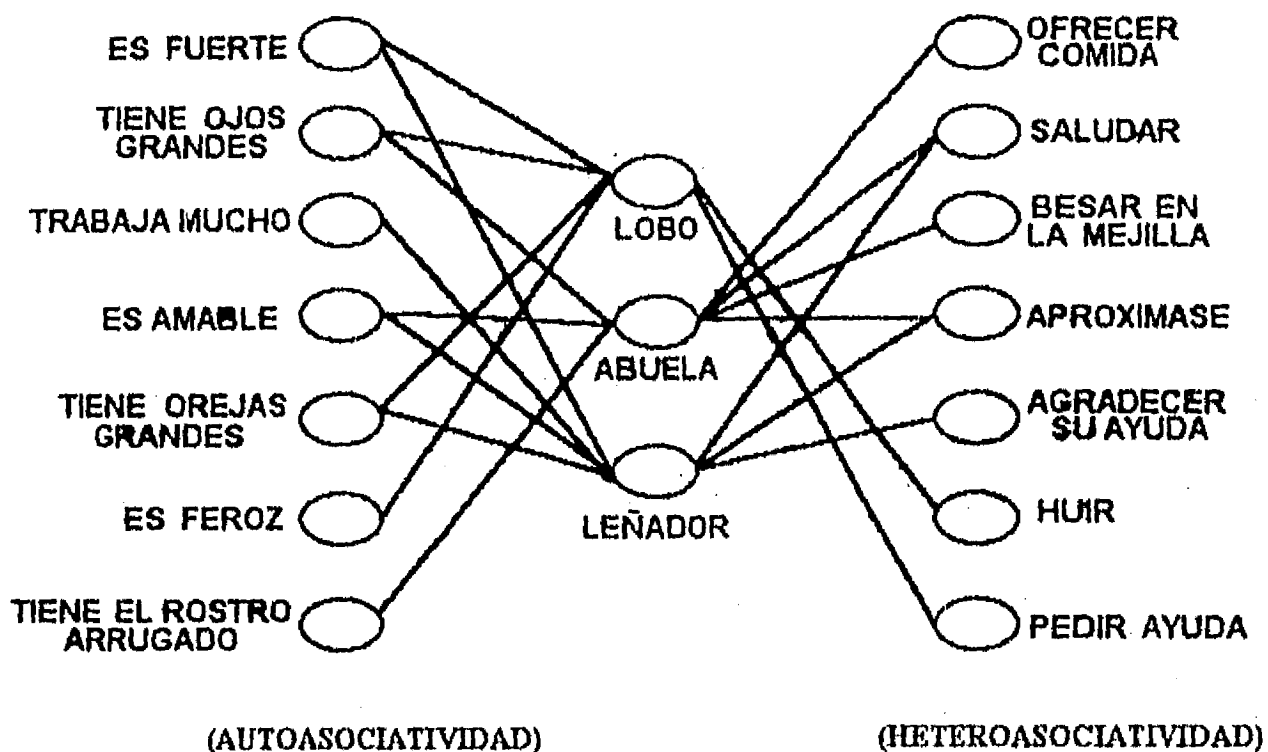
Figura II.16. Los ajustes en los pesos se hacen para igualar un patrón.

Por su parte, el recuerdo se refiere a cómo procesa la red un estímulo que es presentado, para crear una respuesta. El recuerdo por lo común es parte integral de un proceso de aprendizaje, tal como ocurre cuando una respuesta deseada de la red se compara con la salida que muestra y se generan señales de error, acordes a las diferencias que se observan.

Cierto es que las personas no son videograboras que reproducen con toda exactitud lo que han capturado. Las personas tienden a olvidar, distorsionar, omitir, agregar, etc. lo que recuperan de la información adquirida. Este hecho indica que el humano transforma o elabora la información. Una red no opera así: *no olvida*. Cuanto el nivel de error se reduce, el de recuerdo se eleva. Sobre esta base puede inferirse que el *aprendizaje deseado está ocurriendo*.

Al hablar de *entrenamiento*, ya sea de una persona, de un sistema o una red, simplemente expresamos que se trata de una fase en que se expone al aprendiz a la información para que la revise, la relacione con base en algún criterio de organización, por ejemplo formando categorías, la almacene mediante representaciones apropiadas, y que la pueda recuperar cuando se requiera.

Pero, *¿cuál sería un aprendizaje deseable?* Aquí hay por lo menos dos opciones, una de tipo *autoasociativo* y otra de naturaleza *heteroasociativa*. Si lo que se desea como producto es que la red atienda a las características de una información dada, entonces la entrada y la salida de la información tienen que ser iguales, es decir, procederá por igualación, reconocimiento o identificación.



SI... ENTONCES

Figura II.18. Un neurocomputador es un reconocedor de rasgos.

Según se advierte en este arreglo, un prototipo podrá aprender a identificar a los personajes atendiendo diferencialmente a los rasgos que los definen.

Tabla II.1. Ejemplo de una matriz de aprendizaje por autoasociatividad

RASGOS	LOBO	ABUELITA	LEÑADOR
es fuerte	1	0	1
es amable	0	1	1
es feroz	1	0	0
trabaja mucho	0	0	1
tiene ojos grandes	1	1	0
tiene orejas grandes	1	0	1
tiene el rostro arrugado	0	1	0

De acuerdo con esta matriz, todas las columnas en que aparece un 1 son rasgos que definen a un personaje y dado que en algunos aspectos se parecen será necesario atender a los rasgos diferenciales, para poder identificarlos. Con matrices similares un neurocomputador puede establecer los rasgos de interés y determinar los patrones (en cuáles casos hay un 1 y

en cuáles hay un 0) que se asocian a un modelo, un perfil de desempeño, una tendencia, etc.

Una vez que un personaje es identificado, el aprendizaje *heteroasociativo* permitirá que Caperucita muestre una respuesta apropiada, como se muestra en la siguiente tabla.

Tabla II.2 Ejemplo de una matriz de aprendizaje por heteroasociatividad.

PERSONAJES			RESPUESTA ADECUADA
LOBO	ABUELITA	LENADOR	
0	1	0	ofrecer comida
0	1	1	aproximarse
0	1	1	saludar
0	1	0	besar en la mejilla
1	0	0	pedir ayuda
1	0	0	huir
0	0	1	agradecer su ayuda

Como se muestra en la matriz, el aprendizaje *heteroasociativo* (indicado en las celdillas con 1), permite generar respuestas apropiadas dependiendo de las características percibidas por *autoasociatividad* en cada personaje. La conducta (acercarse, huir, ofrecer comida, saludar) variará en cada caso. Volviendo al funcionamiento de un neurocomputador, cuando se presentan matrices como la presentada, el prototipo no realiza acciones pero sí indica las que son adecuadas.

Otro aspecto importante es que para diseñar un entrenamiento se puede partir de un modelo, buscando reproducir sus características definitorias. A esto se le llama *aprendizaje supervisado* y para lograr este propósito es común que se realicen numerosos ensayos y que gradualmente se vayan eliminando errores, hasta que como producto final se iguale al modelo. O bien, cuando no se dispone de un modelo, puede diseñarse una estrategia de *aprendizaje no supervisado* y dejar que el procesamiento vaya definiendo los patrones inherentes a los datos, ayudando así a descubrir tendencias, perfiles, etc. La diferencia es clave, ya que esto definirá la operación básica de una red. El caso de la *autoasociatividad* es típicamente supervisado.

De acuerdo con Rosenberg y Blalock (op. cit.), las reglas de aprendizaje más frecuentemente usadas en neurocomputación son: *aprendizaje competitivo*, que fue desarrollada por Kohonen (1989), *Delta* (Wildrow y Hoff, 1960) basada en el procedimiento *LMS* (least-mean-squared) y la *Delta Generalizada*, que se aplica al procedimiento de retropropagación (Rumelhart, Hinton y Williams, 1986).

Cuando durante el entrenamiento se produce un error, la red no procede al ajuste de un componente particular que esté fallando. Lo hace en forma global regresando la información a todos los elementos, gracias a un mecanismo básico de retropropagación del error. Y cuando los ensayos para lograr el resultado deseado muestran avances se aplica una regla de aprendizaje, por ejemplo la *Delta*, para fortalecer los aciertos. De esta manera, junto con otros mecanismos, se va regulando la propagación de la información.

¿De dónde provienen los valores que determinan los pesos de las señales que reciben las neurodas? Esto lo define el usuario mediante matrices de datos que pueden estar formuladas en formatos binarios o de otros tipos. Los modelos matemáticos que subyacen a la operación realizan cálculos y ajustes necesarios para que se cumplan los procesos de entrenamiento, aprendizaje y recuerdo exitosamente.



Figura II.19. El usuario define lo que un neurocomputador debe aprender.

Debe señalarse que un sistema automatizado o *maquinístico* no sustituye al experto humano, al menos en la actualidad, pero permite reproducir su expertez y hacerla extensiva a otros medios, por ejemplo en los que no hay uno disponible. Al mismo tiempo, permite aumentar sus capacidades considerablemente. Cabe decir que en el presente estudio fue posible reconocer tres patrones mediante la lectura de 1500 renglones con 8 columnas de datos, que contenían los resultados de 1500 sujetos en ocho evaluaciones, en unos cuantos minutos.

Cuando este trabajo de análisis lo efectúa una persona, la búsqueda tiende a ser lenta, con alto riesgo de cometer errores y tiempo de procesamiento indefinido. Por esa razón, puede afirmarse con fundamento que el reconocer patrones no requiere de la participación de una persona para ser exitoso. Aún más, en entornos propios de la Inteligencia Artificial, se ha ido ganando terreno en la identificación de patrones, gracias a su capacidad para descubrir y mostrar relaciones emergentes que no habían sido consideradas por el investigador (White, 1989), en entornos de aprendizaje no supervisado.

Sin embargo, un sistema de neurocomputación sí requiere ser alimentado por un experto humano, incorporando los datos pertinentes

mediante matrices que se construyen para comunicarse con el prototipo. En el presente caso el punto de inicio fue realizar análisis psicométricos cuidadosos de las respuestas de los sujetos evaluados, como base para el procesamiento automatizado.

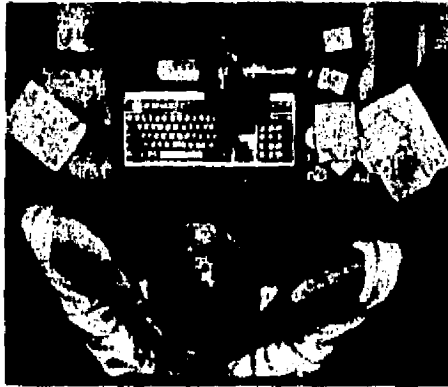


Figura 11.20. *Un neurocomputador no se autoalimenta, debe hacerlo el usuario.*

La información que se prepara para el procesamiento no es una tarea que a la fecha, al menos en nuestro medio, realice una máquina y, por su importancia, tiene mucho peso en los resultados.

Desde otra perspectiva, Simon (1979) y Haugeland (1993), señalan que el que un sistema sea artificial o simulado no lo invalida por ese simple hecho. Por ejemplo, si no se contara con sistemas automatizados que permitieran seguir y simular en computadora las posibles trayectorias de un huracán, el nivel de riesgo aumentaría muy considerablemente. En otro sentido, la tecnología que ha hecho posible generar insulina sintética no demerita su alto valor médico.

Cabe señalar que la operación de un neurocomputador no se basa en un programa con rutinas definidas por el usuario, como ocurre en otros sistemas de diagnóstico automatizado. Lo que ocurre es que la información generada en el entrenamiento está difundida en toda la red, por lo que cada componente tiene una parte de ella, en vez de estar guardada en algún almacén central o periférico.

Por esto se dice que la operación es en paralelo y es global. Si una parte de la red se pierde no ocurre algo irremediable, como cuando al recitar un poema olvidamos un verso. Las demás unidades conservan su capacidad de activación y el funcionamiento continúa. De hecho, esta característica se asemeja mucho a la que tiene lugar en el procesamiento humano y la recuperación no literal, pues podemos operar con información incompleta, difusa y a veces errónea.

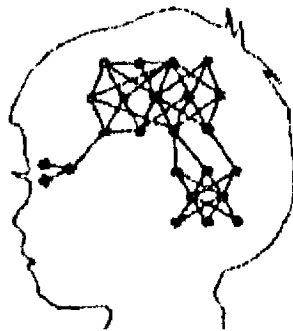


Figura II.21. El conocimiento está difundido en toda la red.

En conclusión, el conocimiento es más una función de la arquitectura o estructura de la red, que del contenido de una dirección particular y consiste en un estado general o condición de equilibrio.

El comportamiento de las redes neurodadas en las funciones de aprendizaje y de recuerdo de un patrón puede aclararse con un ejemplo de sus mecanismos básicos.

Si se asignara el valor 1 a un foco encendido y 0 a uno apagado, se podría decir que, en un momento dado, en una casa el jardín, la cochera, la sala y el comedor tendrían un indicador de 1 al estar encendidos los focos y el resto, por estar apagados, tendrían valor 0. Esta forma de codificar la información es útil para presentar en forma breve y precisa un conjunto de datos, por ejemplo, para tomar decisiones.

Si todos los lugares de la casa tuvieran en este momento valor 1 sin que fuera necesario, se podría describir este hecho como un perfil o patrón de uso inadecuado de la energía eléctrica. Para los fines de este ejemplo, una matriz con los datos del uso de la electricidad diría:

JARDIN COCHERA SALA COMEDOR RECÁMARA = PATRON INDESEABLE
 1 1 1 1 1

Este patrón, basado en la identificación del estado de una serie de indicadores bien definidos, proporciona una especie de diagnóstico a partir del cual se pueden generar acciones que sean consecuentes con lo observado. Por ejemplo, apagar las luces innecesarias:

JARDIN COCHERA SALA COMEDOR RECÁMARA = PATRON DESEABLE
 0 0 1 1 0

El punto que se desea destacar es que al reunir una serie de indicadores que permiten describir una situación o un estado de cosas, se está definiendo un patrón que puede ser más o menos característico de dicha situación o estado, como se ilustra a continuación con un caso

hipotético de los resultados finales de tres alumnos típicos del primer año de secundaria:

Alumno 1

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGÍA	GEOGRAFÍA	FÍSICA	DEPORTE
CALIFICACION:	9.5	10	9.5	9.5	10	10

Alumno 2

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGÍA	GEOGRAFÍA	FÍSICA	DEPORTE
CALIFICACION:	8	7	7	7	7	7

Alumno 3

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGÍA	GEOGRAFÍA	FÍSICA	DEPORTE
CALIFICACION:	5	4	3	5	4	6

Estos datos del desempeño de tres alumnos parecen sugerir tres diferentes perfiles, atendiendo a los valores que se muestran en cada una de las materias. Si codificamos las respuestas en una escala de 0 a 1, atendiendo al nivel más alto (1) y al más bajo (0), considerando además un valor medio (.5), la *matriz* podría quedar de la siguiente forma:

Alumno 1

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGIA	GEOGRAFIA	FISICA	DEPORTE
CALIFICACION:	1	1	1	1	1	1

Alumno 2

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGIA	GEOGRAFIA	FISICA	DEPORTE
CALIFICACION:	.5	.5	.5	.5	.5	.5

Alumno 3

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGIA	GEOGRAFIA	FISICA	DEPORTE
CALIFICACION:	0	0	0	0	0	0

Si a estos resultados se les añadiera un rótulo que distinguiera cada perfil del desempeño escolar en términos de su nivel en las calificaciones, se podría desarrollar una matriz como la que se muestra a continuación.

Alumno 1

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGIA	GEOGRAFIA	FISICA	DEPORTE
CALIFICACION:	1	1	1	1	1	1

NIVEL: ALTO

Alumno 2

MATERIAS:	HISTORIA	ESPAÑOL	BIOLOGIA	GEOGRAFIA	FISICA	DEPORTE
CALIFICACION:	.5	.5	.5	.5	.5	.5

NIVEL: MEDIO

Alumno 3

MATERIAS: HISTORIA ESPAÑOL BIOLOGIA GEOGRAFIA FISICA DEPORTE

CALIFICACION: 0 0 0 0 0 0

NIVEL: BAJO

Sin duda, una pregunta pertinente sería: ¿para qué se está haciendo todo lo anterior? La respuesta es: con este procedimiento simple se está construyendo una forma precisa y sistemática de comunicación con un neurocomputador. ¿Para qué? Para entrenarlo a que aprenda a reconocer estos patrones. ¿Para qué, si ya se conocen los resultados? Para que cuando se presenten los datos de otros estudiantes (por ejemplo calificaciones), el prototipo indique automáticamente a cuál categoría corresponde cada uno de ellos.

Si la primera etapa (entrenamiento) se logra con un margen de error no mayor a 0.10, la tarea de reconocimiento se producirá, muy probablemente, con valor de error igual a cero. Este nivel es el apropiado porque, si por ejemplo se fijara a $=.001$, lo cual es posible, se estaría restando al prototipo su capacidad para generalizar, ya que se trataría de un aprendizaje de casos únicos.

Desde luego, los resultados dependerán de por lo menos dos cosas: a) del éxito que se tenga en construir un neurocomputador adecuado y confiable y b) de la calidad de la información con que se alimente al neurocomputador.

Esta sencilla descripción refleja de manera simplificada el procedimiento que se siguió en la presente investigación, para generar las matrices fuente para el entrenamiento y la prueba; muestra el enorme potencial de aplicaciones de este desarrollo tecnológico en una extensa gama de problemas relacionados con la categorización, el reconocimiento de patrones y la secuencia de acciones, que pueden ser abordados en el campo psicológico.

Por último, puede señalarse que el nombre de *neurocomputación* asignado a este método de análisis puede ocultar el valor generalizado de su utilidad, ya que, como ha sido planteado, consiste en un sistema poderoso para reconocer patrones automatizadamente, identificar y reconocer rasgos, tendencias, etc. que no está circunscrito a alguna disciplina en particular, por lo que su aplicación se ha vuelto común en materia de ciencias biomédicas, comunicaciones e ingeniería, entre otras.

En la siguiente sección se expondrán algunos antecedentes y aplicaciones del desarrollo neurocomputacional, que ilustran aspectos interesantes para la psicología.

b) Algunos antecedentes en el campo de la Neurocomputación relevantes para la Psicología

El campo de la neurocomputación no es nuevo: tuvo sus inicios en los años 40's. En esta década, el neurobiólogo McCulloch y el estadístico Pitts en 1943 publicaron el trabajo titulado "A Logical Calculus of Ideas Imminent in Nervous Activity" que fue un antecedente importante para el desarrollo de tres campos importantes: las primeras computadoras digitales (John von Neumann); Inteligencia Artificial (Marvin Minsky) en términos de *inteligencia macroscópica*, y los trabajos de Frank Rosenblatt (perceptron) sobre computación y óptica (reconocimiento de patrones).

En diferentes momentos, este campo se ha denominado de distintas formas: sistemas neuroartificiales, conexionistas, adaptativos, procesador distribuido en paralelo, circuitos de decisión colectiva y, como se les conoce actualmente, neurocomputadores.

Tal vez el antecedente más significativo para este campo ocurrió hace 40 años. En 1956 Minsky, McCarthy, Rochester y Shannon, organizaron la *Primera Conferencia Internacional de Inteligencia Artificial*. En este evento se discutió acerca del uso potencial de las computadoras en la simulación de cada uno de los aspectos del aprendizaje o cualquier rasgo de inteligencia. A partir de la década de los años ochenta, señala Caudill (1989) la neurocomputación tuvo un resurgimiento importante, al valorar las ventajas de esta tecnología frente a sus competidores: cercanos los sistemas expertos gobernados por reglas y algoritmos poco flexibles.

Este resurgimiento fue precedido por una investigación extensa en los grandes centros internacionales de desarrollo y aplicación científica y tecnológica, a la vez que los resultados impactaban el interés de disciplinas muy diversas. Como reseñan Lawrence y Luedeking (1991), se le empezó a relacionar con estudios de problemas financieros, predicción del clima, toma de decisiones empresariales, modelamiento conductual, identificación y reconocimiento de patrones, diagnóstico clínico, investigación biomédica, robótica, lógica, lenguaje. Las autoras antes citadas señalan que las aplicaciones neurocomputacionales más significativas no han estado ligadas a un campo disciplinario particular, porque se les utiliza lo mismo en el Centro Médico de la Universidad de Nueva York, que en la NASA, TRW, servicios de defensa, Ford, General Electric y, por supuesto, en numerosas universidades que son ampliamente reconocidas.

También en el campo de la psicología se han desarrollado diversos modelos útiles, es decir, que no son sólo experimentales. Por ejemplo, los que se relacionan con los procesos de *aprendizaje* (Hinton, 1987; Gluck y Bower, 1988; Kohonen, 1989), *interpretación del lenguaje natural* (Waltz y Pollack, 1988), *inteligencia* (Grossberg, 1978), *lectura en voz alta* (Sejnowski

y Rosenberg, 1987), *diagnóstico, detección y reconocimiento de patrones* (Rumelhart y McClelland, 1982; Pao, 1989), *representación del conocimiento* (Caudill, 1989), *modelos de memoria* (Feldman, 1981; Anderson, 1983; McClelland y Rumelhart, 1985).

En México, hasta donde el autor conoce, en el *campo educativo* sólo se cuenta con los trabajos de López, Castañeda, Pineda y Orduña (1992) y de Castañeda (1993), mientras que en *neurociencias* se han hecho numerosas aplicaciones. Por ejemplo: Villanueva (1990), Espinoza (1990), Carrillo (1990), Rojas y Garda (1990), además de un número importante de artículos publicados por Negrete y colaboradores (1980, 1988, 1990, 1992).

Las redes neurodales pueden ser clasificadas arbitrariamente por su topología, modelo y algoritmo de aprendizaje, mediante dos subdivisiones principales: los modelos de redes que basan su operación en la realimentación y las que lo hacen mediante la alimentación hacia adelante.

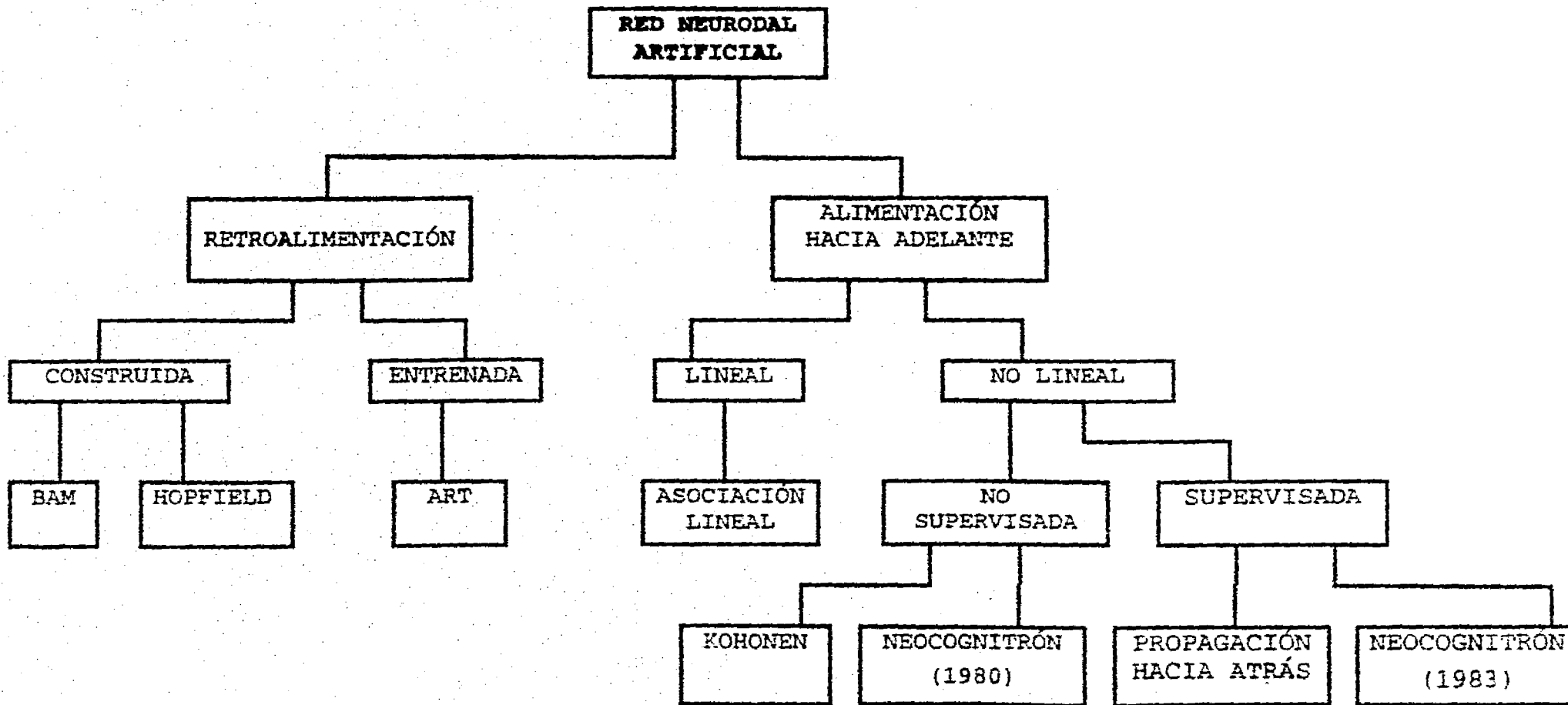
La figura que se presenta en la siguiente página muestra modalidades de redes neurocomputacionales y sus derivaciones, que de acuerdo con Lawrence (1990), son las más representativas en este campo. La división más general se da entre las redes que basan su procesamiento en la retroalimentación y las que lo hacen mediante la alimentación hacia adelante, tomando como referencia el flujo de la información durante el entrenamiento. De cada división se desprenden grupos o familias de arquitecturas de redes.

Para ilustrar algunas aplicaciones del diseño e instrumentación de las redes neurodales, se describirán en forma sumaria algunos trabajos cercanos al ámbito psicológico, en que se ha seguido el modelo de la neurocomputación.

En 1959 Widrow, de Stanford, desarrolló un elemento adaptativo lineal llamado *Adaline* (adaptive linear neuron) basado en elementos simples como las neuronas. La *Adaline* y una variante de dos capas, la *Madaline* (Multiple Adaline) fueron usados para una variedad de aplicaciones, incluyendo *reconocimiento del habla, reconocimiento de caracteres y control adaptativo*. Más tarde, la *adalina* fue modificada para producir una salida continua, más que discreta.

Anderson (1977, 1983), de la Universidad de Brown, enfocó su trabajo pionero al desarrollo de un modelo, llamado *asociador lineal*, enfocado al almacenamiento en memoria, recuperación y reconocimiento. Este *asociador* es un modelo distribuido en paralelo, fundamentado en el principio de Hebb (1949), de que la conexión entre dos elementos, como las neuronas, se fortalece cada vez que éstas son activadas.

Figura 11.22. Clasificación de las redes neurocomputacionales.



El trabajo de Kohonen (1977, 1989), de la Universidad Técnica de Helsinki, Finlandia, inició un avance fundamental en la investigación acerca del *aprendizaje adaptativo y la memoria asociativa*. A él se debe el análisis y la descripción de una cantidad amplia de reglas de adaptación locales. Se trata de reglas cuyos pesos son modificados en una manera que depende sólo del valor de los pesos previos y los valores pos y pre sinápticos. Las memorias definidas por estas reglas de adaptación incluyen matrices de memorias de autocorrelación que tienen como casos especiales el asociador lineal y el modelo de Anderson.

Otra contribución de Kohonen (1989) es el *aprendizaje competitivo*, en el cual las neurodas compiten entre sí para responder a un estímulo de entrada y el ganador se adapta para responder más fuertemente a esos estímulos (como en el condicionamiento clásico, en que un perro aprende a responder al sonido de una campana que antes era inefectivo). Dicho aprendizaje es no supervisado en el sentido en que la organización interna de la red es gobernada por los estímulos de entrada.

Grossberg (1978, 1988, 1989), también ha desarrollado un trabajo muy importante, usando datos neurológicos para construir modelos, en particular de los mecanismos de *percepción y memoria*. La mayoría de sus reglas formales y postulados fueron derivados de estudios fisiológicos.

En el sistema de Grossberg las unidades básicas pueden tomar sólo un valor de activación real entre un mínimo y un máximo. La función de salida es una función sigmoidea, que es una función umbral típica en neurocomputación del valor de activación de la unidad básica. Una clase de redes que ha sido estudiada extensamente cae bajo el título de *Teoría de la Resonancia Adaptativa* y de ellos se han destacado propiedades que reflejan sus acuerdos con el modelamiento fisiológico.

El modelo de *activación interactiva de percepción de letras* de McClelland y Rumelhart (1981), ilustra cómo una red de conexiones es una forma natural para expresar computaciones en que las representaciones de bajo y alto nivel deben ser computadas simultáneamente para que puedan influirse mutuamente. Las neurodas del modelo interactúan y se organizan para el procesamiento en tres niveles de: rasgos, letras y palabras.

Hopfield, de Caltech (1982), presentó un trabajo en la Academia Nacional de Ciencias de los EEUU bajo el título Modelo Hopfield o red asociativa entre modalidades, acerca del *sistema olfatorio*, que se basó en la clase general de redes considerada por Grossberg. Este modelo representa la operación de la neuroda como una operación de umbral e ilustra la memoria como información almacenada en las interconexiones entre dos unidades neuronales. También ilustró y modeló la habilidad del cerebro para evocar respuestas de muchos lugares en respuesta a un estímulo, con base

en un sistema neurodal que asocia simultáneamente información de muchos sitios de almacenamiento para una entrada dada.

Gluck y Bower (1988) de la Universidad de Stanford, construyeron redes adaptativas simples para explorar algunos aspectos del *aprendizaje humano*. Diseñaron tres experimentos relacionados con la clasificación de pacientes de acuerdo a síntomas que tenían diferentes niveles de correlación con dos enfermedades.

El arreglo neurocomputacional siguió una regla de aprendizaje basada en el método LMS (least-mean-squares) asociado al modelo Rescorla-Wagner, que se ha utilizado en estudios de condicionamiento pavloviano. La red predijo los resultados de esos experimentos, en comparación con algunos modelos de aprendizaje competitivo.

Para proceder a la elección de los patrones, emplearon el modelo LMS para valorar las proporciones de elección observadas para cada uno de los 15 patrones de síntomas, obtuvieron los pesos sumados y convirtieron esa activación en la asignación de un paciente a una categoría de enfermedad rara.

El mejor valor de ajuste cercano a 0, estimado por el *método de los mínimos cuadrados* con los datos de las proporciones observadas, fue muy satisfactorio ya que la correlación entre las proporciones de elección observadas y predichas por la red fue de .94, con una discrepancia media absoluta de .07, lo que sugirió un ajuste razonable del modelo de datos. Sin embargo, la χ^2 de 46.2 (14) $p < .001$ indicó la existencia de desviaciones significativas, atribuidas a efectos de las muestras relativamente pequeñas con que se trabajó, que contenían errores estándar grandes.

Waltz y Pollack (1986) desarrollaron un sistema de procesamiento de lenguaje natural con fuentes de conocimiento modular interactivo. El sistema ofrece el descubrimiento de una variedad de fenómenos lingüísticos y permite la prueba de un sin número de hipótesis.

La interpretación del lenguaje se basó en la activación de la red que fue generada dinámicamente a partir del input, el contexto reciente y el conocimiento de largo término. En este estudio, el comportamiento inicial de la red era ambiguo e inestable, o bien, incidía en una sola interpretación. Al avanzar en el entrenamiento, el modelo fue exitoso en el procesamiento en paralelo para la representación del contexto y la primacía de conceptos, con el apoyo de ejemplos ilustrativos de la influencia contextual sobre la interpretación del significado y el procesamiento semántico de oraciones, entre otros aspectos.

Waltz y Pollack (op. cit) trabajaron en varios niveles: acceso lexical, manejo sintáctico, autonomía en los juicios, integración de información y manejo contextual. Su modelo neurocomputacional se aplicó a la activación

distribuida y la inhibición lateral, de manera que sólo una parte de la red permanece activa en un momento dado, lo que favorece que algunos procesos puedan ser simulados realísticamente. En vez de que los sistemas de acción sean holísticos o simultáneos, se hacen flexibles y secuenciales.

Rosenberg y Blellock (1988) de la Universidad de Princeton y el Laboratorio de Inteligencia Artificial del MIT, respectivamente, trabajaron en la implementación de una red de aprendizaje, desarrollando un sistema poderoso, llamado *Connection Machine*, que tiene una gran capacidad de trabajo en procesamiento paralelo real, gracias a sus capacidades físicas diseñadas para aplicaciones específicas.

Su aplicación se basó en el sistema *NetTalk*, construido en 1987 por Sejnowski y Rosenberg. Se trata de una red que aprende del mapeo de un texto en inglés la pronunciación de dicho texto. Las unidades que componen a la red fueron arregladas en tres capas totalmente conectadas. La primera de las capas (input) recibe información del texto. La pronunciación es computada por la red y es encodificada como el patrón de actividad de las unidades en la tercera capa (output) de la red.

En este trabajo Rosenberg y Blellock (op. cit.) asumieron que las redes conexionistas son modelos de la cognición que permiten transformar muchas de nuestras nociones de representación del conocimiento y el aprendizaje. Discuten la aplicación de un algoritmo conexionista de aprendizaje (retropropagación del error), como base para la operación de una computadora en paralelo.

El algoritmo de aprendizaje de la retropropagación del error fue hecho en la *Connection Machine*, computadora hipercúbica consistente en más de 65,536 (2^{16}) procesadores de un bit. Las redes contienen más de 65,000 cadenas que pueden ser simuladas a velocidades cercanas al doble de lo que hace una Cray-2 (considerada como super computadora). Usando procesadores *virtuales* pueden operar redes con más de 16 millones de cadenas sin un sacrificio significativo en la velocidad. Se observó que el impedimento mayor para aumentar la velocidad fue el establecimiento de comunicaciones entre dos procesadores y no en cada procesador de sí mismo.

La conducta total del sistema es modificada por el ajuste de los valores a las conexiones que determinan cómo se afectarán entre sí las unidades de procesamiento. Esto ocurre a través de la aplicación repetida de una regla de aprendizaje, que permiten a los sistemas aprender por la experiencia o de los ejemplos. En esta red, las unidades están dispuestas dentro de capas, de tal manera que dos unidades en la misma capa no pueden estar directamente conectadas. Sin embargo puede haber un número arbitrario de capas en la red. La información fluye hacia adentro

desde la primera capa (input) hasta la o las capas intermedias, a la última capa (output).

Rosenberg y Blellock (op. cit.) concluyen que las redes conexionistas masivas en paralelo han marcado un redescubrimiento en el campo de la Inteligencia Artificial y la Ciencia Cognitiva y pueden constituir un paradigma mayor para esos campos. El aprendizaje resultante de la decodificación puede ser analizado para descubrir nuevos hechos acerca del dominio modelado. Las ventajas de las computadoras en paralelo se harán más claras conforme se avance en el desarrollo de este tipo de computación.

Es importante señalar las consideraciones de McClelland, quien en 1988 se dio a la tarea de recabar evidencia psicológica de la utilización de los modelos conexionistas en nuestra área. En su revisión ilustra por ejemplo, cómo el modelamiento puede revelar implicaciones empíricas útiles para revisar algunos principios generales y también para formalizar ciertos principios propuestos en el procesamiento de información. Asume que este marco tiende al descubrimiento de nuevas clases de aplicaciones para los hallazgos básicos. De igual manera, la investigación conduce a dar cuenta de fenómenos contradictorios, dispares o no unificados en el campo y hace clara la relevancia de ciertos tipos de evidencia para formular cuestionamientos básicos de la naturaleza misma del sistema de procesamiento.

McClelland señala que en los estudios sobre *cognición humana* en los años 60's era muy común pensar en los sistemas de procesamiento humano de información, como diseños que se parecían mucho a las computadoras de von Neumann, ya que trabajaban en secuencias de operaciones discretas.

En esta década la memoria era vista como un conjunto de almacenes separados; los procesos complejos eran entendidos como diagramas de flujo que solo especificaban secuencias de pasos controlados por un ejecutivo rígido. Las teorías del procesamiento del lenguaje y de su adquisición, de la solución de problemas, de la comprensión del texto y de la representación del conocimiento y del desarrollo de la inteligencia, entre otros, estaban fuertemente impregnadas por la investigación psicológica inspirada en los diseños *von Neumann*, como en los casos de Anderson y Bower (1973) y Anderson (1983).

Hacia el inicio de los 80's un creciente grupo de investigadores aceleró la exploración de otros marcos de referencia, buscando que fueran apropiados para la exploración de diversos procesos cognoscitivos. Así, se empezó a considerar que la metáfora mente-cerebro, útil para una descripción de la macroestructura de la cognición, no era suficiente para el estudio de la micro estructura de dichos procesos (McClelland, 1988).

La alternativa que comenzó a explorarse fueron los modelos conexionistas, de acuerdo a la nomenclatura dada por Feldman (1981), muy relacionada con la psicología experimental cognoscitiva. McClelland (1988) menciona trabajos de Dell; Elman y McClelland; Gluck y Bower; Estes, entre otros, como evidencias de avances significativos en la nueva perspectiva. Los tópicos abordados en estos trabajos fueron: representación activa del conocimiento, procesamiento de información, conocimiento, aprendizaje, medio ambiente y pensamiento.

En referencia especial a la microestructura de la cognición, McClelland (1988) cita los trabajos de Rumelhart, Smolensky y Hinton, que se orientaron a la simulación de procesos complejos como la solución de problemas, representación temporal e imaginaria, que tuvieron efectos en la concepción de los aspectos micro y macro cognoscitivos, incluso de sus aspectos estructurales.

Debe señalarse que no obstante los avances importantes en este campo, algunos autores indican la existencia de aspectos que no han sido cabalmente consolidados. Uno de ellos se refiere a la relación entre conexionismo y modelos neurocomputacionales, que no ha sido muy clara en la perspectiva psicológica.

Por ejemplo, la psicología cognoscitiva no ha mostrado mayor interés por los mecanismos de actividad nerviosa, pero sí por los que están vinculados con la investigación experimental y la caracterización de aspectos del procesamiento humano de información, en que incide la neurocomputación. En tanto que los modelos conductuales parecen inclinarse más por las descripciones funcionales que por el trabajo de experimentación neurofisiológica. Sin embargo, comparte con la aproximación neurocomputacional un interés marcado por los procesos de modelamiento (aprender del ejemplo) y moldeamiento (igualar un patrón).

Hay también críticas de importancia hacia los modelos conexionistas y, por inclusión, de los neurocomputacionales. Por ejemplo, Massaro (1988) plantea que el conexionismo carece de un marco teórico amplio y que consiste más bien en implementaciones de un sistema de procesamiento que podría ser caracterizado de una manera más abstracta. En este punto resulta difícil determinar cuál debe ser el punto óptimo de descripción de un fenómeno cognoscitivo o conductual. Es cierto que muchos procesos en estudio no muestran aún una claridad suficiente, que permita derivar explicaciones en distintos niveles descriptivos.

Massaro (op. cit.) señala que hay diferencias no sólo en las concepciones que se tienen del conexionismo, sino que también es un tanto difuso el concepto en sí mismo. Este autor toma como referencia particular el análisis del desempeño lingüístico presentado en descripciones del procesamiento de información, generadas en el contexto conexionista. Esta

situación de controversia recuerda la ocurrida entre Chomsky y Skinner, dado que sus marcos de referencia eran ostensiblemente distintos.

Estas observaciones reproducen la confrontación que muchas veces se ha dado en nuestra disciplina: la de quienes sostienen que hay mucha teoría y pocos datos y la de quienes afirman que hay muchos datos y poca teoría.

En lo que se refiere al presente trabajo, el autor coincide con la afirmación de Lawrence y Luedeking (1991) y Caudill, (1995), quienes realizan trabajo de investigación y han hecho aplicaciones en el campo de la neurocomputación: *se está apenas en la infancia de un dominio de conocimiento amplio, rico y diverso que forma parte de un esfuerzo interdisciplinario (Inteligencia Artificial), para dar respuesta a necesidades más prácticas que teóricas, sin demeritar el trabajo importante que se ha hecho desde otras aproximaciones, como la lógica y la filosofía.*

Es evidente que aún las aplicaciones neurocomputacionales más complejas, son tan sólo una aproximación microscópica que no reflejan ni de la manera más pálida la estructura y funcionamiento del cerebro, del sistema nervioso o de un conjunto organizado de mecanismos celulares, que incluso en su forma más simple hacen posible pasar de la sensación a la respuesta motora. Pero debe también quedar claro que no ese es el propósito de los trabajos realizados en este entorno.

Cabe también precisar que *simulación* no es sinónimo de *emulación*, ya que esta última pretende reproducir con detalle y fidelidad algunos mecanismos y funciones. Por ejemplo, creando un modelo de fibra nerviosa, un ojo robótico y algunos aparatos que basan su operación en la concepción cibernética para realizar actividades muy especializadas. En tanto que la simulación se interesa por la función, no por la estructura o la forma para realizar actividades diseñadas para cumplir ciertos objetivos. El que esta cualidad los haga más o menos inteligentes tampoco es una cuestión crucial.

Para concluir esta sección, cabe citar fragmentos de dos trabajos mencionados en ocasión de la inauguración de la Cátedra Extraordinaria Enrique O. Aragón en la Facultad de Psicología de la UNAM (8 de enero de 1989), cuyo tema central fue: *Inteligencia Artificial en Psicología*. Ambos artículos aparecieron en el volumen XX, número 12, agosto de 1966, de la Revista de la Universidad de México (páginas 11-13 y 30-32).

En el primero de estos trabajos el biomatemático Negrete (1966) habla de Cibernética y sistemas dinámicos en biología; en el segundo, el psicólogo Díaz-Guerrero trata acerca de Cibernética, psicología y ciencias del comportamiento.

Negrete: "Son en sí extraordinariamente interesantes los conocimientos que se han obtenido del análisis cibernético de los sistemas dinámicos en biología y muy fructífero, a no dudarlo, el estímulo que han producido para promover investigación fisiológica, pero, ¿no será también cierto que estos mecanismos diseñados a través de cientos de miles de años de experiencia de la naturaleza puedan ser usados para mejorar el conocimiento y la tecnología de la raza humana?"

¿No será posible que estos mecanismos se apliquen a la construcción de mejores predictores de señales estocásticas, al diseño de mejores computadoras o de nuevas álgebras? Parte de estos conocimientos biológicos han sido ya utilizados, otros están propuestos para su utilización, pero otros, quizá los más trascendentes, esperan del trabajo coordinado de investigadores que sin el prejuicio generado por su formación básica laboren en la tierra de nadie de la ciencia: la cibernética." (páginas 11-13).

Díaz-Guerrero: "Para Wiener, esta capacidad de retro-alimentación de máquinas y seres vivientes constituye una de las grandes armas en la batalla de los seres vivientes en contra del eterno enemigo: la entropía, la desorganización de la energía.

Pero esta batalla bien pudiera terminar en una condición de jaque con la naturaleza, de no existir la posibilidad, que dentro de la psicología llamamos de aprendizaje, y que dentro del modelo de las máquinas sería la utilización de la retroalimentación, pero en tal forma que previos aspectos de retroalimentación vayan dejando huellas en una memoria de la máquina, y que estas huellas puedan ser utilizadas en el futuro para resolver problemas semejantes a los que anteriormente han sido resueltos, o bien para determinar reglas de conducta ante diversas situaciones ambientales. La posibilidad de aprendizaje en hombres y máquinas viene a ser uno de los elementos más importantes en la batalla que libra el hombre ante la tendencia natural del universo al desorden, a la desorganización y a la confusión, es decir, a la entropía.

Así pues, cuando la información que procede de la actividad de las máquinas o de los organismos es capaz de cambiar el método general o el patrón de la actividad, hemos llegado a aquel proceso que podemos llamar aprendizaje.

Es precisamente en la psicología del aprendizaje, en la neurofisiología del aprendizaje, en donde la cibernética se une a estas ciencias, y donde, por primera vez, se puede observar que los autómatas y las computadoras pueden quizás servir como modelos que traten de semejar o de simular los procesos que se desarrollan en el sistema nervioso, o bien las leyes que los rigen y la naturaleza misma de los fenómenos psicológicos". (páginas 30-32).

Estas dos exposiciones, realizadas hace 30 años, muestran visiones muy claras de acontecimientos que han tenido lugar en tiempos recientes. Llamen la atención dos hechos que sin duda han influido de manera contundente en los avances logrados: la necesidad de unir esfuerzos interdisciplinariamente y la importancia de la investigación centrada en el aprendizaje. Estos dos aspectos son el corazón de la Inteligencia Artificial en general y del procesamiento paralelo neurocomputacional en lo particular.

En la siguiente sección se describirá cómo fue abordado en este trabajo el diseño del prototipo neurocomputacional para el reconocimiento de patrones de desempeño académico.

c) Arquitectura y operación del prototipo de Neurocomputador.

Como fue expuesto al inicio del presente capítulo un diseñador de redes no especifica el algoritmo que ejecutará cada unidad de procesamiento o *neuroda*, como lo haría el programador de una máquina tradicional. Lo que se requiere planear detallada y específicamente es cómo será su *arquitectura*, lo cual implica esencialmente el definir:

- ◆ los valores de entrada
- ◆ el número de capas que se asuman indispensables
- ◆ los valores de salida que mostrarán los avances en la operación de la red
- ◆ el valor meta o patrón a alcanzar en la salida de la red
- ◆ las interconexiones entre las capas de procesamiento
- ◆ la función de transferencia
- ◆ las reglas de entrenamiento de la red
- ◆ el nivel de convergencia
- ◆ el modelo de neurocomputación

Sin embargo, tal vez los puntos más críticos sean: 1) decidir para qué se requiere un neurocomputador, es decir, a qué obedecerá su diseño y construcción y 2) de qué información se dispondrá para alimentar la operación y probar la red, dando por sentado que el paso uno indicará al usuario a qué la aplicará.

Estos dos puntos están muy relacionados. Por principio de cuentas y como suele suceder con todo, un procedimiento puede ser ideal para unas cosas, para otras será más o menos útil y para otras será muy inapropiado. Si no se tiene claridad para qué se desea y para qué sirve un neurocomputador, la tarea será errática. Igual puede ocurrir con los datos: pueden parecer muy interesantes, pero ser poco apropiados para que un neurocomputador los procese.

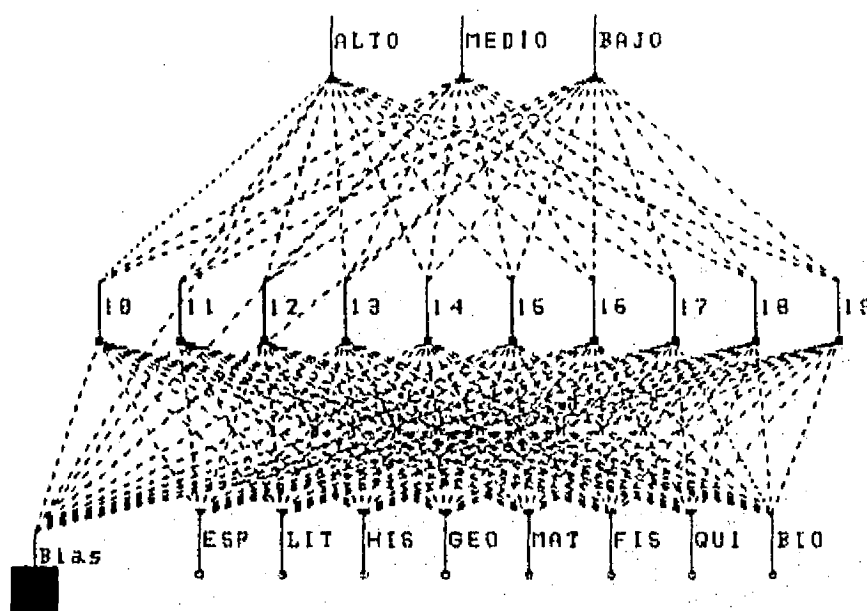
En el presente caso, la respuesta a la primera pregunta ya ha sido expresada en páginas anteriores: aplicar esta tecnología al reconocimiento de patrones diferenciales del desempeño académico. Para lograr esta meta se tomó la decisión de recorrer todas las fases que permiten explorar sus características: diseñar la aplicación, generar un prototipo, probar su operación y aplicarlo al diagnóstico psicoeducativo.

El tipo de información seleccionado para construir matrices con los datos de entrada (entrenamiento) y definir el modelo del reconocimiento *meta* (salida), fue muy similar a la que generan los estudiantes al resolver exámenes, tratando en todo lo posible que fueran datos crudos, para evitar artificialidad en el procesamiento.

Las categorías inherentes a los patrones de respuesta son los que comúnmente se emplean para identificar los niveles de desempeño: alto, medio y bajo. El contenido de los exámenes y el formato de los reactivos también es similar al que se utiliza para evaluar conocimientos y habilidades académicas.

Figura II.23. Arquitectura de la red del prototipo neurocomputacional.

PROTOTIPO DE NEUROCOMPUTADOR BACKPROPAGATION-DELTA



La red generada en el entorno de *NeuralNetworks* para representar en este trabajo las capas neuronales y sus conexiones, se ilustra en la figura II.23. Puede observarse que: a) hay ocho entradas que se corresponden con las asignaturas evaluadas en las subpruebas; b) en la porción intermedia hay 10 unidades con funciones de enlace e integración y c) la salida está

formada por tres neurodas, una para cada categoría de estudiante. Hay además un valor constante (*bias*) igual a 1, necesario para normalizar los pesos de la red.

Especificaciones de la arquitectura del prototipo.

Valores de entrada: las sumas de aciertos en cada una de las ocho materias evaluadas

Número de capas: tres (entrada-enlace-salida). La primera tuvo ocho neurodas, la segunda diez y la tercera tres.

Valores de salida: pesos de la matriz que indicarán el nivel de error entre la respuesta esperada y la observada.

Valor meta o patrón a alcanzar en la salida de la red: tres categorías de desempeño. En un rango de 0 a 1, el valor cercano a 0 se interpreta como patrón de sujetos con desempeño bajo; el valor cercano a 0.5 se interpreta como patrón de sujetos con desempeño medio y el valor cercano a 1 se interpreta como patrón de sujetos con desempeño alto.

Interconexiones entre las capas de procesamiento: sólo inter-capas.

Función de transferencia: sigmoidea

Reglas de entrenamiento: alimentación rápida hacia adelante y retropropagación del error.

Nivel de convergencia: el límite de error aceptable en el entrenamiento y la prueba, que por lo común son valores entre 0.10 y .001.

Modelo de neurocomputación: heteroasociativo con aprendizaje supervisado.

Procedimiento seguido para la operación del prototipo.

Para la operación de la red se siguió el procedimiento que se describe a continuación.

- 1. Definir cómo deberá operar el procesamiento.** Esto incluye determinar en dónde se ubican los archivos con los datos de entrada, en cuáles archivos se guardarán los datos del procesamiento, si se trata de una fase de entrenamiento o de prueba, si se aplicará una red entrenada para el reconocimiento para alguna aplicación, cuál será el nivel de tolerancia al error, etc.

- 2. Alimentar la red con información que servirá para el entrenamiento y prueba.** Esto puede hacerse por dos medios. El más común es presentar una matriz con las especificaciones, en un formato que

pueda ser leído y procesado por el prototipo. Otra forma es introducir desde el teclado de la computadora los datos que lo alimentarán.

3. Monitorear el proceso. Una vez que se inicia la operación del sistema y dependiendo del programa computacional que se utilice, en la pantalla se podrá monitorear el procesamiento y, por lo general, se irán registrando en archivos de datos los cambios en los valores, hasta que se obtiene el resultado deseado, todo lo cual puede ser impreso para su inspección. Se llama *netware* genéricamente al programa computacional que define la estructura de una red y las reglas para resolver un problema específico.

El monitoreo de la operación y análisis de los resultados permitirán tomar decisiones acerca de si se está obteniendo lo planeado o si se requieren ajustes en algún sentido, siendo esto último casi la regla.

4. Verificar si la red concluyó la fase de entrenamiento. En términos generales, se asume que el neurocomputador habrá aprendido cuando pueda generar los patrones de activación correspondientes a los valores de las entradas dentro del margen de error que se haya determinado. La manera como lo haga dependerá del paradigma de aprendizaje seguido.

5. Probar la calidad del neurocomputador. El siguiente paso será probar la red, utilizando datos para los cuales se sabe la respuesta, de manera que se podrá comprobar la precisión de la operación. Esto puede hacerse también introduciendo los datos desde el teclado de la computadora o mediante listados o matrices de prueba. En esta fase los patrones a reconocer son iguales a los del entrenamiento, excepto que no se trata de los mismos objetos o sujetos. Por ejemplo, si se van a utilizar los datos de 1000 sujetos, pueden seleccionarse al azar 200 casos y reservarlos para la fase de prueba.

6. Aplicar el neurocomputador para las realizar aplicaciones previstas. Si los resultados son satisfactorios (valores de error cercanos a cero), entonces se considera que la red está entrenada y puede procederse a la utilización planeada, con datos que no han sido expuestos antes al neurocomputador.

7. Actualizar o depurar los datos de entrada a la red. Este paso permitirá que el neurocomputador mantenga un funcionamiento óptimo, ya sea desechando datos que pierdan vigencia o bien agregando datos nuevos para tener una actualización permanente. De aquí se pasaría nuevamente al *punto 2*.

Esta somera descripción de un procedimiento general refleja con precisión los pasos seguidos en el presente estudio. En la práctica implica tomar decisiones frecuentes y con diferentes niveles de complejidad. No

obstante, la experiencia va haciendo cada vez más ágil el procedimiento. Si, como es propio del novato, se procede por ensayo y error, la tarea puede sentirse como árdua y, eventualmente, con una cierta dosis de frustración. Esto puede ocurrir, entre otras cosas, porque al principio no es fácil saber qué está haciendo la red durante su operación; o bien, para encontrar claridad en los datos y valores que se muestran en pantalla y listados, por ejemplo, para determinar qué nivel de entrenamiento tiene la red.

Una situación importante es el nivel de previsión que se haya hecho, por ejemplo para disponer de varios niveles de análisis de los datos de interés, asumiendo que un problema puede ser abordado con aproximaciones alternativas que permitan explotar óptimamente las capacidades del neurocomputador.

En este trabajo, al construir las matrices con los datos se tuvo en cuenta la posibilidad de poder efectuar varios niveles de análisis, desde el individual hasta el de la población completa. Y por razones prácticas se decidió tomar como base para el reconocimiento de patrones la información de cada sujeto en dos sentidos: sumar el total de aciertos en cada subprueba y obtener el total de aciertos. Esto obedeció a la intención de no atomizar demasiado la información, por ejemplo, haciendo el análisis por reactivo (como se hizo en la parte psicométrica), ya que de una u otra forma los totales tendrían el mismo peso en el patrón del alumno, indistintamente de si hubiera o no congruencia en la combinación de aciertos para alcanzar dichos totales.

Las propiedades que permitieron aprender al neurocomputador y muestran las principales características de las redes neurales (Lawrence, 1991) fueron:

1. La función de activación: se trata de una función que especifica lo que deben hacer las neurodas con las señales que reciben, después de que los pesos han tenido su efecto. Una vez que están dentro de la neuroda, las señales pesadas son sumadas y combinadas con las del estado previo de la neuroda y comparadas contra un valor umbral. Si el umbral es alcanzado, se genera una señal de salida (igual al valor de activación de la neuroda). La función puede ser determinística (variables con valores bien delimitados), como en el caso del neurocomputador construido, o estocástica (valores probabilísticos o tomados al azar), que no se aplica al modelo seleccionado.

El valor de activación es la suma de los valores de entrada ya pesados, en un momento dado: $a_i(t)$. Este puede ser continuo (aceptar valores fraccionales) o discreto. Lo más común son rangos entre 0 y 1 o entre -1 y +1 para indicar un estado activado o no activado. Como fue expuesto en la primera sección, debe tenerse en cuenta que las conexiones de la red tienen pesos, que indican qué tanta fuerza debe tener una señal

para ser pasada a la red, por lo que sus efectos pueden ser excitatorios o inhibitorios.

2. Adaptabilidad: es la capacidad de la red para modificar su operación en respuesta a las condiciones cambiantes. Cuatro procesos contribuyen: **(a)** el aprendizaje, **(b)** auto-organización, **(c)** generalización y **(d)** entrenamiento.

3. Retropropagación: es un método de aprendizaje supervisado que se utilizó para la operación del neurocomputador, en el cual una señal de error es enviada atrás a través de la red, alterando los pesos que lo originan, para prevenir que vuelva a ocurrir el mismo error. Las redes que lo emplean son de alimentación rápida y usan la propagación hacia atrás como una regla de aprendizaje que corresponde a una variante de la regla Delta, por lo que también se denomina "Regla Delta Generalizada". La cantidad que está siendo minimizada es la media del cuadrado del error, que se enuncia como sigue:

$$\frac{1}{N} \sum_{i=1}^N (d_i - o_i)^2$$

Donde d_i es el resultado deseado de la neuroda i , o_i es la salida actual y N es el número de neurodas.

La regla Delta de aprendizaje establece que si hay una diferencia entre el patrón de la salida actual y la del patrón deseado en el entrenamiento, entonces los pesos serán cambiados para reducir la diferencia. La cantidad de cambio en los pesos es considerada con base en el error en la salida, los valores de entrada y la tasa de aprendizaje. La formalización de la Regla Delta se presenta a continuación:

Delta $w_{ij} = \eta(t_i - a_i) o_j$, donde η es la tasa de aprendizaje, t_i es el patrón de enseñanza, a_i es el valor de activación para la neuroda receptora, o_j es la salida de la neuroda que está enviando la señal. También se conoce como LMS (Least Mean Squared Rule).

4. Nivel de convergencia: valor que indica a la red que debe permanecer estable, una vez que hizo todos los ajustes para alcanzar una meta. Una red que ya aprendió es una red estable. Este valor es importante porque marca un límite de tolerancia al error y si este fuera cero en el entrenamiento, la red en vez de aprender memorizaría hechos únicos, afectando su capacidad de generalización.

5. Memoria distribuida: significa que un dato dado no está almacenado en alguna dirección específica, sino que la información está esparcida a través del sistema en paralelo altamente interconectado.

6. Alimentación rápida: se refiere a que las neurodas toman las señales de entrada sólo de la capa anterior y envían sus salidas sólo a la siguiente capa. Por esta razón, las neurodas en las capas del prototipo no están conectadas una con otra. Esto permite que la red pueda computar un resultado con rapidez.

7. Generalización: capacidad de la red para producir una salida deseada cuando se le presentan valores de entrada similares a los que sirvieron de base para entrenarla.

8. Aprendizaje: proceso por el cual la red modifica sus pesos en respuesta a información externa.

9. Tasa de aprendizaje: un factor para escalar todas las correcciones mientras ocurre el aprendizaje, aumentando la velocidad de convergencia.

10. Regla de aprendizaje: ecuación que indica a la red cómo modificar los valores de los pesos de las conexiones. Como ya fue señalado, en el presente caso se aplicó la *Regla Delta Generalizada* ya que el neurocomputador tiene tres capas y la regla Delta simple se aplica cuando sólo se tienen dos capas.

Las reglas de aprendizaje son variaciones de la *Regla de Hebb*, que asume que los cambios en los pesos son proporcionales al producto de las actividades sinápticas tanto de las neurodas que envían las señales como de las que reciben. Se expresa así: $\Delta w_{ij} = \eta a_i o_j$, donde η es la tasa de aprendizaje, a_i es el valor de activación para la neuroda receptora y o_j es la salida de la neuroda que envía la señal.

11. Asociación lineal: las neurodas responden a cambios en las entradas al cambiar la tasa de activación de sus salidas. La red mapea entradas similares con las salidas similares, favoreciendo la capacidad para generalizar.

12. Mapeo: transformación de los valores de entrada a una representación interna en la red, de manera que las relaciones topológicas de las entradas sean similares a las que integran la representación.

13. Memorización: capacidad de la red para producir la salida deseada cuando un dato de entrada fue visto durante el entrenamiento.

14. Reconocimiento de patrones: capacidad para identificar un conjunto de datos previamente aprendido, aún cuando el patrón tenga alguna distorsión o tenga *ruido* (los datos sean imprecisos).

15. Autoajuste: capacidad de la red para adaptarse a los cambios en las entradas al modificar sus procesos internos.

16. Auto-organización: capacidad de las neurodas para ser entrenadas. Esto se logra usualmente por la modificación de los pesos sinápticos individuales en respuesta a los cambios en las entradas. Todos

los pesos de la neuroda son modificados de acuerdo a las reglas de aprendizaje.

17. Función de transferencia: define cómo serán los valores de activación para las señales de salida de la neuroda. En el presente estudio la función es *sigmoidea*, ya que tiene un límite de saturación alto (1), uno bajo (0) y rangos proporcionales (0.05) entre ambos. Esta función es 0 cuando el valor de la activación es marcadamente negativo y de 1 cuando es marcadamente positivo, con un alisamiento de transición entre ambos. Su nombre se debe a que su representación gráfica tiene forma de S.

18. Función de sumación: equivale a la función de activación, por la cual las señales son sumadas y comparadas contra el valor umbral interno.

19. Función de umbral: es una función de transferencia específica con una salida de todo o nada. Si el valor de entrada es mayor que algún valor fijo (el umbral) la neuroda tendrá una salida con valor de 1. De otra forma, el valor será de 0. Por lo tanto, el valor de entrada debe alcanzar un cierto nivel para poder contribuir a un nuevo estado de la red.

20. Aprendizaje supervisado: información a la red acerca de si la salida que produce es o no correcta (modelo de entrenamiento). La salida observada es comparada con la deseada y los pesos se ajustan para reducir la diferencia.

Como pudo observarse en los 20 puntos expuestos, el diseño y operación de un entorno neurocomputacional se basa en principios bien establecidos y de hecho cada arreglo es específico a la tarea que se espera que cumpla. Esta tarea puede facilitarse cuando se sigue la estructura de modelos generales que aportan componentes tan importantes como las reglas de aprendizaje, las estrategias de propagación de la información y las funciones de transferencia.

Si bien no puede calificarse a esta empresa como fácil o sencilla, tampoco es en extremo difícil o compleja. Un balance de tipo costo-beneficio muestra que el provecho de su utilización es una razón muy convincente para incorporar a la neurocomputación como parte de las herramientas útiles para apoyar a la investigación y a la práctica psicológicas.

CAPITULO III

Problema de Investigación y Diseño del Estudio

En las instituciones educativas la evaluación de los estudiantes ha tendido a limitarse a acreditar el conocimiento, soslayando dos funciones importantes que pueden desprenderse de su utilización: profundizar el conocimiento de variables con alto valor predictivo del desempeño académico y realimentar el proceso de enseñanza-aprendizaje. Esta tendencia muestra ausencia de sensibilidad hacia la necesidad de conocer con oportunidad y exactitud qué es lo que aprenden en realidad los estudiantes en las aulas y acerca de las habilidades que requieren ser fortalecidas para incrementar las capacidades académicas para adquirir, elaborar, estructurar, recuperar y aplicar conocimiento útil, por ejemplo, para resolver problemas.

La problemática grave a que da lugar esta evaluación insuficiente puede observarse en todos los ámbitos educativos de nuestro país y expresa, en términos simples, que se ignoran aspectos vitales que deberían ser la fuente para promover cambios curriculares, seleccionar los contenidos de estudio, prescribir prácticas remediales, favorecer el desarrollo de habilidades de pensamiento. Aún más, al conjuntarse esta falta de información con otros factores, por ejemplo los que se asocian a numerosos aspectos de las diferencias individuales, la situación se torna densa y más compleja.

Tal estado de cosas hace difícil identificar claramente a qué se deben los niveles bajos en el desempeño y los altos en deserción escolar, así como el poder determinar los patrones diferenciales de los conocimientos y habilidades de los alumnos en las diferentes asignaturas que cursan. Lo único que se observa con nitidez es el resultado pobre en la realización de tareas instruccionales, a veces tan básicas como la comprensión de lectura.

Puede asumirse que no se trata del efecto de factores aislados, sino más bien de la interacción entre ellos. Por ejemplo, señala Bransford (1979) procesos psicológicos como atención y memoria, interactúan con variables inherentes a los materiales, las tareas y los procedimientos de evaluación en los diferentes contextos instruccionales, a los cuales es sensible la estructura cognoscitiva de los sujetos. Al modificarse o manipularse cualquiera de estos componentes, se observan diferencias en el desempeño. A veces una relectura rápida que ayude a reactivar el conocimiento previo es suficiente para marcar la diferencia; en otras, lo único que requiere el alumno es más tiempo para dar su respuesta.

Se trata, en general, de interacciones entre tres estructuras de naturaleza distinta: a) la conceptual que es propia de los diversos dominios de contenido (física, química, matemáticas, entre otras disciplinas), b) la cognoscitiva, que es propia del estudiante; y c) la del entorno instruccional, que controla, articula y determina formas cambiantes de relación entre las estructuras del contenido y las cognoscitivas, formando un todo. Es decir, se conjugan para constituir una base, vale decir un patrón, que interrelaciona elementos de ambas organizaciones, ya sea apoyando o interfiriendo la integración de la información y con ello, el aprendizaje, la comprensión y el recuerdo de lo aprendido.

A la luz de los diversos problemas planteados cabe enfatizar la importancia de realizar evaluaciones cuanti y cualitativas, que permitan identificar y contrastar los patrones de aprendizaje en distintos contextos instruccionales. Se asume que hay rasgos distintivos que tienden a agruparse y que subyacen a la construcción de conocimientos relevantes a la tarea académica, como los del tipo *declarativo*, integrado principalmente por hechos, conceptos y principios; y los instrumentales, como el *procedural* y el *condicional* (Anderson, 1985). Asimismo, que ayuden a esclarecer cómo adquirimos, interrelacionamos, modificamos, estructuramos y aplicamos el conocimiento (Gagné, 1991).

Todo esto bajo la consideración de que también existe una diversidad de contenidos, propósitos y situaciones presentes en los diferentes contextos de medición del aprendizaje (tipos de pruebas) y de tareas (demandas de ejecución requeridas por los ítems), que dificultan valorar cuáles estrategias metodológicas son idóneas para lograr identificar con precisión los patrones más típicos del desempeño académico.

En este contexto, para las universidades y en general para los centros de enseñanza, constituye una prioridad y sin duda también un reto el poder contar con procedimientos, instrumentos y estrategias de evaluación de alta calidad. Por ejemplo, para sustentar un proceso justo, válido y confiable que coadyuve a identificar y seleccionar a quienes deben aceptar en sus aulas.

El interés central en la presente investigación está enmarcado en una parte específica de la problemática planteada: la evaluación para el ingreso a la educación media superior, en particular al nivel de bachillerato.

Esta evaluación puede considerarse especialmente relevante porque es el puente obligado para ingresar a la educación superior y porque arroja información valiosa del inicio de un proceso formativo que tiene como primer paso superar un examen de admisión. Sin duda, es también un punto estratégico para explorar numerosos elementos de las trayectorias de respuesta al examen, que mostrarán estudiantes con desempeños

diferentes. El conocimiento de estos aspectos tiene implicaciones importantes para el proceso educativo.

La aproximación psicológica apropiada para su tratamiento es el estudio experimental con fines de diagnóstico, ya que se pretende mostrar los aspectos relevantes que contribuyen a conformar una situación específica, en este caso los perfiles diferenciales que se asocian al desempeño alto, medio y bajo en la evaluación.

La tesis que se sostiene en esta investigación es que a los desempeños distintos les subyacen patrones de respuesta específicos. Para identificarlos es necesario analizar las respuestas de sujetos de distintos niveles de ejecución, en pruebas que exploren muestras representativas de contenidos, habilidades y tipos de conocimiento pertinentes a las características de los sujetos y a los propósitos de la evaluación. Los patrones que sean identificados podrán enunciarse como perfiles de habilidad, una vez que los aspectos diferenciales estén definidos.

Un aspecto igualmente relevante y complementario de este planteamiento, es que el reconocimiento de los perfiles de interés puede hacerse en forma automatizada mediante modelos computacionales diseñados para operar con rapidez y precisión en el análisis de patrones de respuesta.

La identificación de los patrones puede basarse en los dos parámetros de mayor impacto en la calificación: la habilidad de los sujetos y la dificultad de los ítems de la prueba. Para lograrlo deben emplearse instrumentos balanceados y reactivos calibrados que en conjunto aseguren la precisión de la medición y que por su estructura contribuyan a esclarecer aspectos diferenciales en cada sujeto.

Para instrumentar esta propuesta, en la presente investigación se aplicará una estrategia metodológica mixta, basada en dos contextos teóricos distintos en naturaleza pero complementarios funcionalmente: el análisis psicométrico y el procesamiento neurocomputacional.

El primer paso estará centrado en utilizar materiales elaborados *exprofeso* para examinar los conocimientos y las habilidades de interés, aplicando pruebas construidas por personal académico competente, con el apoyo de técnicas que se desprenden de un modelo de evaluación que conjuga componentes:

a) de valor *teórico*, particularmente acerca de los procesos cognoscitivos representados como habilidades académicas y tipos de conocimiento que interactúan con los contenidos específicos de las asignaturas que se evalúan y que se asume que determinan las diferencias en las respuestas a los reactivos.

b) de valor *psicométrico*, como base para establecer los perfiles pertinentes a partir del análisis de las respuestas.

El segundo paso del estudio experimental es presentar la información a un sistema computacional construido para reconocer patrones, en la modalidad de un neurocomputador, para probar si la tarea se realiza confiable y rápidamente.

El objetivo principal queda definido, entonces, por el interés de identificar patrones diferenciales del desempeño académico de muestras de estudiantes, para proceder después a su reconocimiento automatizado. Para el presente caso se trata de egresados del nivel de secundaria, con base en el análisis de las respuestas en una evaluación similar a las que se aplican para el ingreso al bachillerato.

La vía *psicométrica* está enmarcada en la *Teoría de respuesta de reactivos* abreviada *IRT* siguiendo en especial los modelos de Embretson y colaboradores (1981,1983,1985,1991); y los de Wright y colaboradores (1979, 1982,1991,1994). La segunda línea metodológica surgió del campo interdisciplinario que es fuente del procesamiento neurocomputacional: la *Inteligencia Artificial*, Se aplicarán dos modelos instrumentales: Brainmaker (1991) y Neuralworks (1989).

A continuación se expresan los aspectos que fundamentan la elección del problema de investigación:

- 1) Siendo la *psicometría* un conjunto de herramientas útiles y poderosas al servicio de la psicología, debe profundizarse en la comprensión y manejo de los modelos que muestran mayor suficiencia para el tratamiento de problemas de medición y construcción de escalas en el contexto educativo, como los *IRT*.
- 2) Siendo el diagnóstico psicológico una tarea demandante de capacidades y habilidades especializadas y por ende, susceptible de error e imprecisión, debe prestarse una atención especial a la exploración de alternativas para mejorar su calidad, entre las que figura la simulación neurocomputacional.
- 3) Siendo el aprendizaje un tópico central en el campo psicológico, los desarrollos tecnológicos que basan su operación en este proceso, como los sistemas neurocomputacionales, ameritan que se les preste mayor atención, para conocer y valorar las aportaciones que en otros campos se están generando y, sobre esa base, enriquecer a nuestra disciplina.
- 4) Siendo la psicología cognoscitiva una de las corrientes más fortalecidas por los avances teóricos, metodológicos y tecnológicos de mayor impacto en los últimos 30 años, es muy conveniente acercarse más profunda y propositivamente a los desarrollos que han ocurrido tanto en el contexto internacional de las ciencias cognoscitivas en general, como en el de la *Inteligencia Artificial*, en particular.

5) Siendo la *inteligencia* un tema que la psicología desde siempre ha considerado importante por sus amplias implicaciones, por ejemplo para el contexto educativo, no es claro por qué un tópico innovador tan importante como lo es la inteligencia artificial, ha sido ignorado en forma sistemática en nuestro medio.

En relación a estos cinco puntos, cabe hacer algunas consideraciones que apoyan su contextualización y resaltan su importancia e implicaciones en el marco del presente trabajo de investigación.

Realizar correctamente un diagnóstico psicológico, indistintamente de la modalidad de que se trate, no es una tarea simple, fácil o trivial. Así lo prueban, entre otros muchos casos comunes en nuestro medio, el determinar, por ejemplo, si un niño debe o no recibir educación especial; si un estilo de enseñanza tiene o no efectos positivos y duraderos en el aprendizaje; si los rasgos inestables o los habituales en el comportamiento de un persona lo inclinarán a un deterioro grave de su salud física o mental; si la forma como se educa a los menores tendrá efectos negativos en su vida adulta; si los valores y creencias que se muestrean en un cuestionario tipifican a una cultura y el distinguir si un comportamiento al efectuar un diagnóstico es característico de un experto o de un novato.

Se asume que estas tareas son propias de especialistas que tienen a su favor una experiencia muy amplia en los problemas de su competencia (no se es *experto* en todo), gracias a una formación teórico-conceptual amplia que los dota de un marco de referencia orientador y a las habilidades que entrenan para el reconocimiento de indicadores útiles, todo lo cual permite llegar a conclusiones certeras y tomar decisiones apropiadas.

Las diferencias entre un diagnosticador experto y uno novato pueden plantearse de diferentes maneras. Seguramente la más notable estriba en la calidad del proceso y del producto que uno y otro pueden lograr partiendo de un conjunto de hechos, por ejemplo para determinar un curso de acción o los eventos que tendrían lugar si no se modificaran las condiciones actuales. La forma estereotipada o maneras alternativas que se sigan para representar el problema, para aplicar estrategias de búsqueda e identificación de indicios y para seguir métodos de comprobación de hipótesis, serán factores que también afectarán la calidad de un diagnóstico.

En esencia, un diagnóstico consiste en el reconocimiento de un patrón, al que le subyace una estructura que puede ser inferida o reconocida aplicando estrategias o reglas de razonamiento, por ejemplo: "*si... entonces*", es decir una condición antecedente y una consecuente, que da lugar a una conclusión o una acción. Esta forma de análisis se apoya en la representación del conocimiento de tipo condicional y, en términos de la teoría *ACT* (Anderson, 1983), es parte del aprendizaje procedural.

En esta perspectiva, la base del razonamiento que subyace al diagnóstico consiste en establecer con precisión la primera condición: "si ...", para pasar válidamente a la prescriptiva: "entonces...". La formulación presupone realizar secuencias estratégicas de operaciones o acciones basadas en algoritmos o procedimientos de búsqueda orientadas al reconocimiento de patrones. Estos procedimientos no han podido ser plasmados en una teoría unificada, tal vez por la complejidad que les es inherente.

Negrete, Castañeda y López (1992) señalan que identificar o reconocer patrones corresponde a actividades que está muy relacionadas con la solución de problemas y, por ello, con la experiencia y la pericia humanas. Para explorar algunos de sus componentes más ostensibles, generaron una propuesta orientada a la *formalización de la conducta del perito*, tratando de encontrar los componentes más relevantes que se asocian a la calidad diferencial entre los expertos y los novatos en este proceso.

En la propuesta se asume que el *enunciado de un problema* es el punto de partida indispensable para identificar los indicadores necesarios para formular un diagnóstico. El proceso consiste en una doble tarea: la de reconocimiento de patrones y la de validación de soluciones, en el contexto del dominio particular de que se trate. Este marco permite establecer diferencias significativas entre novatos y expertos en términos de: a) cómo se representan el problema; b) la estrategia que aplican al procedimiento de búsqueda, por ejemplo si formulan hipótesis iniciales que someterán a prueba o si comienzan el proceso reuniendo datos para configurar un patrón; c) la riqueza y complejidad de la estructura de la base de conocimientos disponible, y d) del sistema de validación de las soluciones.

En ese sentido, Castañeda y López (1992) al analizar la perspectiva que ofrecen los modelos cognoscitivos en torno a la solución experta de problemas, señalan que una fuente común de error en la búsqueda de soluciones es la representación defectuosa del problema a resolver y la aplicación inadecuada de procedimientos que los expertos utilizan. Consideran que esta problemática es pertinente no sólo a un interés teórico de la psicología, sino también a la necesidad de mejorar la tecnología derivada de la construcción de *sistemas expertos*, en el campo de la Inteligencia Artificial.

En esta revisión, expresan que un perito *humano* constituye una instanciación total o parcial de una secuencia de conductas expertas; y que un sistema experto representa una instanciación *maquinística* de tal secuencia de conductas.

En estos términos, un experto no se distingue porque se trate de un humano o de una máquina, sino por los factores de calidad asociados a

cada paso de la secuencia y al resultado de todo el proceso. Los pasos o componentes que cabe asumir como críticos, con relación a un patrón de conductas expertas serían: a) reconocimiento de patrones, b) estrategias para buscar soluciones, c) representación del conocimiento y d) aprendizaje.

El reconocimiento de patrones deriva de la capacidad para manejar los recursos de memoria, de la cual dependen, entre otras, las habilidades para manejar configuraciones. Al jugar ajedrez, por ejemplo, si se detecta una configuración, se activa la representación de varios movimientos posibles en las piezas. Este aspecto, interpretado teóricamente como generación de díadas de producciones del tipo *configuración-acción* orientan a las estrategias de búsqueda. Estas estrategias pueden iniciarse con un proceso de tipo serial deductivo y modificarse adaptativamente para permitir que la recuperación e igualación de patrones en la memoria apoye generar procesos de inferencia para construir hipótesis que constituyen representaciones más abstractas y predictivas del método de solución.

Una característica fuertemente asociada a la conducta experta es el gran cúmulo de conocimiento y experiencia que permite establecer con facilidad discriminaciones y generalizaciones rápidas y eficientes. Este rasgo explica la capacidad para reconocer con precisión patrones complejos y para prescribir con rapidez las acciones apropiadas. La extensión, complejidad y profundidad del conocimiento son en buena parte la base de la calidad de la representación. Por ejemplo, los expertos parecen tener una colección de *modelos aplicables* a la solución de problemas en campos especializados de dominio, en que ejercen su pericia. En ellos es notorio que la organización del conocimiento muestre estructuras jerarquizadas, que presentan mucho más niveles que las de los novatos, como señalan Feltovich, Johnson, Moller y Swanson (1984).

El último componente mencionado en cuanto a los cuatro componentes críticos en el patrón de conductas expertas explorado por Negrete, Castañeda y López (op. cit.) se refiere al aprendizaje o, mejor dicho, a la capacidad para aprender de la experiencia. En efecto, el experto tiende a generar nuevas asociaciones y representaciones al percibir patrones recurrentes en los problemas que examina, ligándoles soluciones o métodos de solución más rápidos y precisos, de mayor abstracción, generalidad e inclusividad. El producto resultante es la reestructuración en varios sentidos, por ejemplo, de las reglas de razonamiento, refinando los pasos para llegar a un resultado.

Lo expuesto en párrafos anteriores lleva a una conclusión que conviene destacar: la formulación de un diagnóstico, lejos de ser un arte que obedece a la mayor o menor inspiración o intuición incidental de una persona, es un proceso estructurado y regulado que se inicia con la búsqueda de un patrón a partir del enunciado de un problema, que sigue

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

reglas, por ejemplo de producción, que permiten decidir cuál operador aplicar, cuándo hacerlo y para evaluar si se está alcanzando o no la meta de encontrar una solución.

En suma, se trata de un proceso muy dinámico y complejo que en el caso de los sistemas *expertos maquinísticos* basados en diseños computacionales, señala Negrete (1980, 1988, 1990, 1992), requiere de un conjunto amplio de reglas bien definidas para la prueba de hipótesis y el cumplimiento exitoso de funciones diagnósticas y prescriptivas. A la tecnología empleada para diseñar estos sistemas se le llama *ingeniería del conocimiento*. La operación de sus componentes está dada por una base de conocimientos y el *motor inferencial* que aplica reglas, como las de producción para llegar a conclusiones válidas. Estos sistemas se construyen para reproducir la pericia humana, tanto para fines aplicados como para apoyar la investigación en diversos campos o dominios, como medicina, ingeniería, psicología, biología, etc.

En el caso de los expertos humanos esta alta estructuración es menos observable pero no menos real, ya que se trata de conductas encubiertas difíciles de registrar en el momento de su ocurrencia. Por ejemplo, cómo se construye la representación del problema y cómo se enuncia su formulación son producto de observaciones reiteradas y reflexiones que descansan en mecanismos de análisis, de síntesis, de inferencia, etc. que forman parte de procesos intelectuales complejos.

La formulación de un problema, dice Kerlinger (1975), no es una actividad sencilla que al primer intento sea clara y completa. Si se desea resolver un problema es preciso saber de qué problema se trata; una parte de la solución reside en saber qué es lo que se intenta hacer. Esto se asemeja a la formulación de un diagnóstico porque suele plantearse como una pregunta que interroga acerca de la relación entre dos o más variables. La calidad de una formulación depende de que se le enuncie sin ambigüedades. Otro criterio es que el problema y su formulación permitan su verificación empírica, es decir, que la relación entre las variables pueda medirse de algún modo.

Es común que la relación entre variables de la forma: "si... entonces", al efectuar un diagnóstico se exprese mediante hipótesis que sean susceptibles de someterse a prueba. Sin embargo, éstas suponen que haya una teoría que sustente su formulación. Dicho de otra forma, las hipótesis son los instrumentos de trabajo de la teoría y lo que las hace relevantes son las relaciones que enuncian. La base que las alimenta son observaciones que se guían por y que guían a las construcciones conceptuales acerca de las variables o fenómenos que forman parte de alguna clase o grupo particular que puede ser subdividido en subconjuntos más suscintos y

exactos para hacer más evidentes las relaciones o formas de asociación entre las variables.

El establecimiento de un diagnóstico o la identificación de un patrón puede verse en forma apropiada más como el reconocimiento de formas de relación entre clases de fenómenos, que como la percepción de rasgos aislados. Este punto permite efectuar comparaciones y, con ello, definir vínculos de inclusividad. Aún más, hace posible medir.

Kerlinger (op.cit), señala que las relaciones en las ciencias se dan entre conjuntos o clases de objetos. Lo que define formalmente una relación es que se trate de conjuntos de pares ordenados, constituidos por dos objetos o por un conjunto de dos elementos que aparecen en un orden fijo. Si atendemos al último señalamiento, podremos inferir que dos o más objetos pueden estar relacionados entre sí, pero que el orden en que se relacionan puede variar y, con ello, configurar patrones diferentes.

Para ilustrar este punto puede tomarse un ejemplo del autor citado: un conjunto de niños puede ser evaluado para establecer su puntuación en aprovechamiento y su cociente intelectual. Esto define una relación entre inteligencia y logro escolar. En el momento en que se expresan estas dos medidas, asociadas al niño específico a que corresponden, se está planteando una relación válida. Si dicha relación se ilustrara en forma gráfica, para representar los dos conjuntos de puntuaciones, cada niño estaría definido expresamente por las dos puntuaciones. Lo que haría válida a la relación es que los miembros de cada par aparecieran en el orden determinado por la medición y no al azar. Si el orden se alterara cambiaría el significado de la información.

Otro componente importante señalado por Kerlinger (op. cit.) en la formulación de un problema para su investigación, tiene que ver con la necesidad de contar no sólo con relaciones válidas para poder inferir los factores a que obedece su vinculación, sino con la de establecer reglas de correspondencia entre los miembros de los diferentes dominios, para el planteamiento de *funciones* bien definidas. Se asume que todas las funciones son relaciones, pero no en el sentido inverso. Por ejemplo, los padres pueden tener varios hijos, pero los hijos tienen sólo un padre. Su enunciado formal sigue expresiones del tipo: "si tenemos p , entonces tendremos q ", lo cual puede escribirse como $q = f(p)$ y leerse "q es función de p".

Esta observación nos recuerda que uno de los objetivos que mejor expresa las metas de la psicología y que se refleja en el establecimiento de diagnósticos, es el que se refiere al interés por descubrir relaciones funcionales entre los fenómenos que constituyen su objeto de estudio. Dichas funciones pueden expresarse en formas matemáticas útiles para hacer deducciones y predecir variables dependientes a partir de las

independientes. Cuando se detectan regularidades en las asociaciones, es decir, que muestran un patrón más o menos consistente, nos preguntamos hasta dónde su ocurrencia está obedeciendo a alguna ley y en qué medida se debe al azar. Lo que suele resultar de la búsqueda de respuestas plausibles a esas preguntas es un nivel de probabilidad. La interpretación de la probabilidad tiende a basarse en la frecuencia de ocurrencia de un evento, con un margen de incertidumbre, como ocurre en los diagnósticos.

El problema central de desarrollar y probar procedimientos probabilísticos que se apliquen válidamente al establecimiento de relaciones funcionales entre variables, sin que los resultados dependan de quien mide, de con qué mide y de a quién mide, como condiciones necesarias para transformar las observaciones en mediciones, es bastante viejo y en los esfuerzos por resolverlo podemos encontrar un camino árduo, frustrante y, con frecuencia, cuestionado.

En efecto, encontrar métodos de medición específicos a la evaluación propuesta e independientes de la variación de otras características de los sujetos medidos, o de los instrumentos de medición empleados, es un reto que ha recibido muchas soluciones y en la historia de la medición psicológica podemos encontrar sus rastros en el cambiante contexto de los tiempos

Por otra parte, diseñar una aplicación tecnológica, con capacidad para realizar en forma automatizada tareas de diagnóstico con calidad de experto, es decir, con error = 0, tampoco es una tarea simple, fácil o trivial. El éxito en esta empresa también plantea retos y demanda una conjunción de competencias especializadas.

Un sistema automatizado o *maquinístico*, no sustituye al experto humano, pero sí permite reproducir su expertez y hacerla extensiva a otros medios, por ejemplo, en los que no hay uno disponible. Al mismo tiempo, permite aumentar algunas capacidades en forma considerable, por ejemplo, para procesar grandes volúmenes de información. En el presente estudio, fue posible reconocer tres patrones de respuesta mediante la lectura de 1500 renglones con 8 columnas de datos con los resultados del desempeño de 1500 sujetos en ocho evaluaciones, en unos cuantos minutos.

Cuando este trabajo de análisis lo efectúa una persona, la búsqueda tiende a ser lenta, con alto riesgo de cometer errores y tiempo de procesamiento indefinido. Por esa razón, en este trabajo se afirma con fundamento, que el reconocer patrones no requiere de la participación de una persona para ser exitoso. Aún más, en entornos propios de la Inteligencia Artificial, se ha ido ganando terreno en la identificación de patrones, gracias a su capacidad para descubrir y mostrar relaciones emergentes no consideradas por el investigador (ver, por ejemplo: Lawrence

y Luedeking, 1991; White, 1989; Gluck y Bower, 1988; Waltz y Pollack, 1986).

Sin embargo, el sistema computarizado sí requiere ser alimentado por un experto humano. La información que se prepara para el procesamiento no es una tarea que a la fecha, al menos en nuestro medio, realice una máquina y, por su importancia, tiene mucho peso en los resultados. Esto justifica el partir, como en el presente caso, de análisis psicométricos cuidadosos como base para el procesamiento automatizado.

Como señalan Simon (1978) y Haugeland (1993), el que un sistema sea artificial o "simulado", no lo invalida por ese hecho. Si no se contara con sistemas que permiten seguir y simular en computadora las posibles trayectorias de un huracán, el nivel de riesgo aumentaría muy considerablemente. En otro sentido, la tecnología que ha hecho posible generar insulina sintética no demerita su alto valor médico.

Simon plantea una tesis que impulsó muy importantemente la investigación en diversas disciplinas, como administración, teoría de la organización, economía y psicología: hay ciertos fenómenos que son *artificiales* en un sentido muy específico, como lo es únicamente porque hay un sistema que, por sus objetivos o sus fines, se amolda al medio en que vive. Si los fenómenos naturales tienen en sí un factor de *necesidad*, a causa de su subordinación a la ley natural, los artificiales poseen un factor de *contingencia*, resultado de la maleabilidad que les confiere el medio.

El planteamiento de esta tesis ha tratado de ser sensible a necesidades de distinta naturaleza: sociales, teóricas, metodológicas, que coinciden en señalar la importancia de explorar soluciones tecnológicas que sean ventajosas y significativas para la realización de tareas complejas, como el diagnóstico psicoeducativo, en apoyo a la investigación y a la aplicación profesional.

El estudio se enfoca a un nivel exploratorio un tanto grueso, al tomar como eje el establecimiento de categorías de estudiante partiendo del desempeño en un examen y tomando como indicador el total de aciertos. Sin embargo, otros avances muestran que es posible reconocer en forma automatizada, con relativa facilidad, otros indicadores que encierran un valor predictivo considerable, no contemplados en el objetivo central del presente trabajo, como podrían ser:

- ¿cómo responden los hombres y cómo las mujeres?
- ¿cómo son los patrones de respuesta de quienes estudiaron en escuelas públicas y cómo son los de quienes estuvieron en escuelas privadas?
- ¿cuáles son los patrones diferenciales de respuesta de estudiantes que se ubican en un mismo percentil?

- ¿cuáles son los patrones diferenciales de estudiantes que tienen el mismo promedio de calificaciones en sus cursos anteriores?
- ¿cómo responden los sujetos de diferentes grupos de edad?

En el siguiente capítulo se describe la estrategia metodológica seguida en la realización de este estudio.

Capítulo IV

Método

1. Sujetos.

1500 estudiantes egresados del nivel de enseñanza secundaria, que fueron distribuidos en cinco grupos independientes, con 300 sujetos cada uno.

Para este estudio no se consideró relevante en los sujetos: edad, sexo, escuela en que cursaron los estudios y promedio escolar. La única variable de interés fue el número total de aciertos en la evaluación que se les aplicó en la presente investigación, la cual sirvió para seleccionar a los sujetos.

2. Materiales e Instrumentos

A. Descripción de los materiales e instrumentos usados en la fase psicométrica.

Con base en el modelo de evaluación descrito en el apartado *d* del capítulo 1, se construyó un instrumento objetivo para evaluar conocimientos y habilidades académicas, identificado como *examen muestra*, formado por 120 reactivos con cinco opciones, una correspondió a la respuesta correcta y cuatro a distractores, siguiendo la estructura y contenido descritos por la SEP (1994) y el CENEVAL (1994).

A continuación se señalan los programas de análisis psicométrico que fueron empleados para la identificación de patrones de desempeño :

1. ÍTEMAN (1993), 2. RASCAL (1992) y 3. BIGSTEPS (1994).

Para análisis estadísticos diversos se emplearon los programas: EXCEL 5.0 (1994), SPSS (1992) y SYSTAT (1992).

B. Descripción de los materiales e instrumentos usados en la fase neurocomputacional.

Los materiales básicos para alimentar al neurocomputador fueron 1500 hojas de respuesta obtenidas en la fase psicométrica, diseñadas para lectora óptica. Para la construcción y operación del neurocomputador se emplearon los siguientes programas: NeuralWorks Professional II (1989) y BrainMaker (1991).

3. Procedimiento

Para lograr las metas planteadas en este estudio, se partió de un análisis psicométrico de dos facetas: alumnos y reactivos, aplicado a las respuestas de 1500 estudiantes egresados de secundaria, asignados al azar a cinco grupos independientes de 300 sujetos: 100 de desempeño alto, 100 medio y 100 bajo. De cada subgrupo de 100 se separan al azar 80 sujetos para la etapa de entrenamiento del neurocomputador y 20 para la prueba expresa en cada aplicación, como se indica en las siguientes figuras.

Figura IV.1. Estructura metodológica de la fase psicométrica

Gpos.	n	Niveles de Habilidad			EVALUACION					
		alto	medio	bajo	1	2	3	4	5	
1	300	100	100	100	300					
2	300	100	100	100		300				
3	300	100	100	100			300			
4	300	100	100	100				300		
5	300	100	100	100					300	
N	1500	500	500	500	300	300	300	300	300	

Figura IV.2. Estructura metodológica de la fase neurocomputacional.

Gpos.	n	HABILIDAD				Total	GRUPOS				Total	1	2	3	4	5
		alta	media	baja	alto		medio	bajo								
1	300	80	80	80	240	20	20	20	60		X	X	X	X		
2	300	80	80	80	240	20	20	20	60	X		X	X	X		
3	300	80	80	80	240	20	20	20	60	X	X		X	X		
4	300	80	80	80	240	20	20	20	60	X	X	X		X		
5	300	80	80	80	240	20	20	20	60	X	X	X	X			
N	1500	400	400	400	1200	100	100	100	300	4	4	4	4	4		
					ENTRENAMIENTO	PRUEBA				APLICACIÓN						

A. Descripción del procedimiento en la fase psicométrica.

A partir de la estructura de un examen muestra compuesto por 120 reactivos se derivaron cinco versiones parcialmente equivalentes, en la

medida en que compartían 77 preguntas comunes. El resto (43) fueron diferentes para cada caso.

En conjunto, los reactivos evaluaron los mismos contenidos en las ocho asignaturas seleccionadas, observando niveles de dificultad similares, pero fueron distintos en el número de reactivos relacionados con la medición de conocimientos y habilidades.

Las versiones difirieron entre sí, también, en la colocación de los reactivos dentro de cada subprueba, controlando de esta manera la conducta de copia. Los grupos se identificaron con 1, 2, 3, 4 y 5, en relación a la versión de examen que contestaron.

Si bien la distribución de reactivos por tipos de habilidad y conocimiento no fue igual en cada versión, por contenido sí lo fue. Este último constituye el criterio clave que fue seleccionado para la comparación, por ser el indicador que se observa comúnmente en contextos escolares para evaluar el desempeño académico.

Sin embargo, como fue explicado al describir el *modelo de evaluación* (capítulo II), para alcanzar un nivel óptimo de explicación, no es suficiente con analizar los contenidos que son relevantes en un dominio de conocimiento, sino que también y tal vez más importante, se requiere explorar las habilidades académicas necesarias para su manejo apropiado.

Un aspecto metodológico interesante a destacar es la disponibilidad actual de diversos modelos estadísticos y psicométricos que han sido diseñados para explorar numerosos componentes de los instrumentos de medición, tanto al nivel de la prueba como de los ítems que las forman; mientras que son raros los que apoyan el análisis a fondo de variables que se asocian más a los sujetos que a los instrumentos con que son medidos.

Este énfasis en cierto modo aleja el propósito de la evaluación, ya que antepone el medio al fin. Tal dificultad es especialmente importante e interesante en el presente estudio, ya que al estar orientado a la identificación de patrones de desempeño, debe profundizar en la aplicación de indicadores apropiados a las características de los sujetos, además de mostrar las cualidades de las pruebas que sustentan las evidencias.

Por esas razones, el análisis principal se hará con base en dos paradigmas metodológicos generales *ad hoc*, como se muestra en las dos figuras siguientes, diseñadas para apoyar la exploración e identificación de los patrones de interés.

El primer paradigma enfoca las respuestas a reactivos de diferentes grados de dificultad, distinguiendo así los niveles de habilidad de los sujetos; el segundo ofrecerá una contraparte, al enfocar los niveles de desempeño diferencial en reactivos con distintos grados de dificultad.

Figura IV.3. Paradigma sujeto-reactivo

ÍTEMS	SUJETOS									Aciertos	GRADOS DE DIFICULTAD
	1	2	3	4	5	6	7	8	9		
1	1	0	0	0	0	0	0	0	0	1	DIFÍCIL
2	1	1	0	0	0	0	0	0	0	2	
3	1	1	1	0	0	0	0	0	0	3	
4	1	1	1	1	0	0	0	0	0	4	MEDIO
5	1	1	1	1	1	0	0	0	0	5	
6	1	1	1	1	1	1	0	0	0	6	
7	1	1	1	1	1	1	1	0	0	7	FÁCIL
8	1	1	1	1	1	1	1	1	0	8	
9	1	1	1	1	1	1	1	1	1	9	
Suma	9	8	7	6	5	4	3	2	1	TOTAL	
	ALTO			MEDIO			BAJO			1 = Acierto	
	NIVELES DE HABILIDAD									0 = Error	

Figura IV.4. Paradigma reactivo-sujeto

SUJETOS	REACTIVOS									Aciertos	HABILIDAD
	9	8	7	6	5	4	3	2	1		
1	0	0	0	0	0	0	0	0	1	1	BAJA
2	0	0	0	0	0	0	0	1	1	2	
3	0	0	0	0	0	0	1	1	1	3	
4	0	0	0	0	0	1	1	1	1	4	MEDIA
5	0	0	0	0	1	1	1	1	1	5	
6	0	0	0	1	1	1	1	1	1	6	
7	0	0	1	1	1	1	1	1	1	7	ALTA
8	0	1	1	1	1	1	1	1	1	8	
9	1	1	1	1	1	1	1	1	1	9	
Aciertos	1	2	3	4	5	6	7	8	9	TOTAL	
	DIFÍCIL			MEDIO			FÁCIL			1 = Acierto	
	GRADOS DE DIFICULTAD									0 = Error	

El ordenamiento de sujetos y reactivos es posterior a la aplicación ya que se establece empíricamente, mientras que en la evaluación el orden de los reactivos puede estar dado por una estrategia diferente. En este caso se intercalaron reactivos de distintos grados de dificultad, en vez de asumir una progresión lineal de mayor a menor dificultad.

Los paradigmas presentados plantean una exigencia metodológica básica de validez y confiabilidad de un instrumento, derivada en particular de los modelos *IRT*: debe existir una correspondencia directa entre el *nivel de habilidad* de los sujetos y el *grado de dificultad* de los reactivos.

Dicho requisito demanda que haya congruencia en las respuestas: los reactivos *fáciles* deben *tener* un número de aciertos significativamente *mayor* que los reactivos de dificultad media o alta; el nivel de habilidad de los sujetos deberá ser expuesto por la dificultad de los reactivos. Por ejemplo, los sujetos con desempeño *bajo* deberán responder correcta y mayoritariamente sólo reactivos de dificultad baja. Si la evidencia no lo confirma, se trataría de un instrumento defectuoso que requiere ajuste y calibración de uno o más reactivos.

Los sujetos con desempeño *alto* mostrarán la mayor congruencia, ya que podrán responder correctamente la mayoría de los reactivos y el nivel de aciertos en los ítems de dificultad alta será significativamente mayor que el de los sujetos con habilidades baja o media. Por otra parte, los sujetos con habilidad *media* tendrán un número de aciertos significativamente mayor que los sujetos de habilidad baja y menor que los de habilidad alta. Finalmente, los sujetos de habilidad baja mostrarán patrones difusos y con alta dispersión, pero se igualarán en el límite de aciertos que logren.

Para delimitar las categorías de *habilidad* en el desempeño de acuerdo al número total de aciertos de los sujetos en el examen, se establece el rango de 1 a 40 para el nivel bajo, 41 a 80 para el medio y 81 a 120 para el alto. De manera similar, los reactivos se ordenan por grado de *dificultad*, de manera que los 40 con el número de aciertos más alto demarcan el nivel fácil, los siguientes 40 el medio y los últimos, con menor número de aciertos, el difícil.

Los contenidos seleccionados para construir el instrumento de evaluación dieron lugar a ocho subpruebas, con número de reactivos diferente en dos casos: biología (12), español (24), física (12), geografía (12), historia (12), literatura (12), matemáticas (24) y química (12). Esta distribución se mantuvo igual en todas las versiones.

Las preguntas de examen se presentaron en cuadernillos con instrucciones y ejemplos para responder; se proporcionó a los sujetos hojas para registrar las respuestas. A partir de ellas se elaboraron matrices y cuadros de concentración para los análisis psicométrico y estadístico de los

datos, mediante ordenamientos por niveles de habilidad de los sujetos y por grados de dificultad de los reactivos.

Para el análisis psicométrico se usaron los programas computarizados que fueron mencionados en el apartado anterior. Antes de describir cómo se usaron dichos materiales, cabe plantear algunos puntos que ayudarán a contextualizar su utilidad para los propósitos de esta investigación.

Los programas para el procesamiento psicométrico, basados en el modelo de Rasch y plasmado en los desarrollos *IRT*, se basan en el análisis del ítem para expresar las propiedades de la prueba. El método de asignar la puntuación pesa en forma automática las respuestas a cada reactivo, para estimar la habilidad del sujeto.

Estos sistemas automatizados ofrecen soporte cuantitativo para computar la probabilidad de que un sujeto responda correctamente, con base en las características del ítem y como una función de la habilidad del sujeto. Se les llama *modelos logísticos* porque transforman los puntajes crudos a unidades escaladas en unidades lógicas, con valores entre 0 a 1.

Tales características de los instrumentos y los procedimientos para la medición, convergieron con el propósito de este estudio: identificar el patrón de respuesta diferencial en el desempeño, al observar las congruencias y los desajustes en los aciertos y errores del sujeto, desde diferentes puntos de inclusión del análisis. Este puede centrarse en el resultado total, en los parciales de las subpruebas, en los de los reactivos particulares, en ordenamientos por niveles de habilidad o dificultad, etc., que se sustentan en formas de relación entre lo esperado y lo observado.

Dado que los programas disponibles muestran capacidades diferentes en el análisis psicométrico de los reactivos, se aplicó un programa que sigue el análisis de tipo tradicional (ÍTEMAN, 1993) y dos que se basa en modelos Rasch (RASCAL, 1992 y BIGSTEPS, 1994). Más que buscar contrastar entre ellos sus ventajas y desventajas, se valoró su complementariedad. A continuación se describe cómo fueron aplicados en todas las versiones de examen.

1. *ÍTEMAN* (1993). Se utilizó para obtener información que es usual en un análisis de tipo *tradicional*: a) número de reactivo en la secuencia original de la prueba; b) proporción/porcentaje de aciertos en el reactivo; c) índice de discriminación del reactivo; d) correlación reactivo-total biserial o punto biserial, entre las respuestas correctas en el reactivo y el puntaje total en la escala a que el ítem fue asignado; e) la proporción de respuestas en cada opción del reactivo; f) la proporción de sujetos en los grupos bajo y alto que respondieron correctamente, y g) el coeficiente de correlación punto biserial entre el total de respuestas en cada opción y en la opción correcta.

Se aplicó también para obtener un sumario estadístico de: media aritmética de aciertos, varianza y desviación estándar, skewness de la distribución, kurtosis, puntajes mínimo y máximo, mediana, valor de confiabilidad Alpha de la homogeneidad de la escala y error estándar de la medición.

2. *RASCAL* (1992) se empleó para hacer análisis tipo Rasch de dos parámetros (discriminación y dificultad) obteniendo: a) puntajes escalados y dificultades escaladas de los reactivos, con estimaciones centradas, por separado, en habilidad y dificultad, basadas en transformaciones lineales, en un margen de + 3 a -3 ; b) errores estándar asintóticos asociados a los parámetros estimados del reactivo; c) conversión de puntajes crudos a indicadores de habilidad expresados en la máxima probabilidad de acertar un reactivo, mostrando los errores estándar de la medición, la frecuencia de respuestas, percentiles y los puntajes escalados; d) un mapa de la relación reactivos-personas, representando gráficamente la distribución de los estimados de la dificultad del reactivo apareados con la distribución de la verdadera habilidad de los sujetos, y e) mapa de la función de respuestas a la prueba, conocido como curva de información de la prueba.

3. *BIGSTEPS* (1994). Se aplicó para obtener indicadores de los desajustes de las personas y los reactivos al modelo, determinar las distribuciones de respuesta más probables y generar varios sumarios estadísticos. Permite analizar patrones de respuesta que muestran incongruencias. Por ejemplo, cuando un sujeto con nivel de habilidad bajo obtiene respuestas correctas que no son consistentes con su patrón. O bien, si no ocurre una respuesta correcta que cabría esperar, al mostrarse un nivel de habilidad alta. Los ordenamientos de los datos por el programa enriquecen las observaciones del comportamiento de los sujetos en las pruebas. Esta información es necesaria para la calibración de los instrumentos de evaluación y el análisis de los patrones de respuesta.

Otros procesamientos, propiamente estadísticos, como correlación múltiple, análisis de varianza y de regresión, se apoyaron en los programas: EXCEL 5.0 (1994), SYSTAT (1992) y SPSS (1992).

B. Descripción del procedimiento en la fase neurocomputacional

El primer paso en esta etapa fue generar matrices binarias con los datos de los sujetos de cada grupo y el segundo consistió en hacer simulaciones del procesamiento, diseñando varias estructuras de redes con diferentes números de neuronas en la capa intermedia, estimando en cada caso el ajuste al modelo deseado.

Un propósito implícito en este estudio era mantener en lo posible el paralelismo con la etapa psicométrica, por lo que fue importante tratar de

utilizar las mismas matrices fuente de los análisis psicométrico y estadístico (puntajes crudos de los sujetos en cada una de las materias evaluadas). Al observar que el entorno NeuralWorks (1989) era útil para generar redes neurocomputacionales, pero en su procesamiento no mostraba un buen ajuste a la estructura de los datos, se optó por utilizar el programa BM (BrainMaker, 1991), que ofrecía una plataforma computacional más apropiada tanto en la definición de los datos de entrada, como en las características de operación. Sobre esta base se efectuó la construcción y prueba del prototipo de neurocomputador.

El prototipo del neurocomputador siguió el algoritmo *retropropagación del error*. La tasa de aprendizaje se fijó en 1.0, con tolerancia de 0.10 y 0.05. En cuanto a su arquitectura, el prototipo tuvo 8 neurodas de entrada (una para cada asignatura evaluada), 10 intermedias (enlace entre entrada y salida, con funciones de integración) y tres de salida (una para cada categoría de habilidad: alta, media, baja). Para la operación del prototipo se usaron matrices apropiadas a las etapas previstas: a) entrenamiento, b) prueba y c) aplicación.

Debe tenerse presente que expresiones como: recuerdo, aprendizaje y entrenamiento, en el contexto neurocomputacional no se refieren a operaciones que realizan los sujetos, sino a las que efectúa el prototipo diseñado para el reconocimiento de patrones. Los únicos datos relacionados con los sujetos se refieren a las puntuaciones que alcanzaron en el examen.

La etapa de entrenamiento consistió en que el neurocomputador leyera y procesara la matriz con los resultados de cada una de las ocho materias evaluadas. Los datos correspondieron a 240 sujetos seleccionados al azar (80 de desempeño alto, 80 medio y 80 bajo), de los 300 iniciales. Para cada sujeto se indicó en la matriz su categoría de desempeño, en atención al número total de aciertos en el examen. El entrenamiento concluyó cuando el neurocomputador mostró un nivel cero de error en la identificación de los patrones. Esta operación se hizo por separado para cada grupo.

La etapa de prueba difirió de la de entrenamiento en que la matriz que leyó el neurocomputador contenía sólo los resultados de 60 sujetos seleccionados al azar, 20 de cada nivel de desempeño, del grupo de 300 que respondió la misma versión de examen, sin indicar su categoría. Los datos de los 60 sujetos seleccionados para esta etapa fueron distintos a los que se usaron para el entrenamiento. La prueba consistió en que el prototipo indicara para cada sujeto su categoría de desempeño, de acuerdo a los patrones que aprendió en la fase de entrenamiento.

La última etapa, de aplicación, fue igual a la de prueba, excepto que la tarea de reconocimiento se realizó con los resultados de sujetos distintos a los que se emplearon para entrenamiento y prueba, sin indicar su categoría de desempeño. El propósito de esta etapa fue obtener evidencia de la medida en que el neuro computador podía identificar correctamente los patrones de respuesta, cuando se trataba de sujetos que compartían algunas características con los que se tomaron como base para el entrenamiento. En otras palabras, se trató de una prueba para explorar la capacidad de generalización y confiabilidad del prototipo.

Por ejemplo, se utilizaron los resultados de 240 sujetos del grupo 1 (versión 1 del examen), para entrenar al neurocomputador (lectura de los datos), es decir, para que aprendiera los patrones de respuesta asociados al nivel de desempeño. Cuando se constató que el error en el reconocimiento era cercano a 0, se presentaron los datos de 60 sujetos del mismo grupo 1, sin indicar la categoría de desempeño. Finalmente, se evaluó en pruebas por separado si se estaban identificando con la precisión requerida los patrones de los grupos 2, 3, 4 y 5.

Para concluir este capítulo, cabe enfatizar que el arreglo metodológico de corte experimental adoptado para realizar la investigación, pone en el primer plano la importancia de proceder sistemáticamente y en lo posible con un control óptimo de las condiciones en que se obtiene evidencia, en la realización de estudios diagnósticos.

CAPITULO V

Resultados

Introducción

La presentación de los resultados se organizará considerando que en el estudio se aplicaron dos enfoques metodológicos: psicométrico (apartado A) y neurocomputacional (apartado B), por lo que los datos serán expuestos en dos apartados temáticos, haciendo subdivisiones que faciliten su revisión. Para contextualizar la descripción de los resultados primero se mostrará la estructura de los *grupos de sujetos* y de los *instrumentos* utilizados, poniendo interés especial en las pruebas de su equivalencia.

Los resultados darán cuenta de los siguientes indicadores del desempeño: 1. sujetos (población total y subdivisión en cinco grupos); 2. prueba aplicada (datos totales de cada versión); 3. aciertos en las subpruebas (resultados en cada asignatura) y 4. aciertos en los tipos de conocimiento y de habilidad académica.

Cuando sea pertinente se establecerá la relación entre nivel de *habilidad* (alto, medio, bajo) de los sujetos y grado de *dificultad* de los reactivos (fácil, medio, difícil) para contrastar los patrones de respuesta y definir los perfiles diferenciales aplicando esos dos parámetros.

La segunda porción de este capítulo, abordará la aproximación *neurocomputacional* y expondrá los datos que se obtuvieron en las fases de entrenamiento y prueba para cada grupo, establecerá contrastes entre éstos y mostrará el perfil general de la población evaluada, tomando como punto de comparación los resultados de la porción psicométrica.

A.1 Resultados del análisis estadístico

A.1.1. Estructura de los grupos y de los instrumentos de medición.

La figura V.1 indica la distribución homogénea de los sujetos, considerando los tres niveles de habilidad que fueron planeados para el estudio. Cada uno de los cinco grupos quedó integrado por 300 sujetos, por lo que la muestra total investigada fue de 1500 sujetos.

La figura V.2 describe la estructura homogénea de las cinco versiones de examen, atendiendo al número de reactivos de los tres grados de dificultad que fueron incluidos. Cada versión está compuesta por 120 reactivos distribuidos en tres subgrupos, con 40 ítems en cada caso.

Figura V.1. Estructura de los cinco grupos.

Grupos	n	Niveles de Desempeño			V1	V2	V3	V4	V5
		alto	medio	bajo					
1	300	100	100	100	300				
2	300	100	100	100		300			
3	300	100	100	100			300		
4	300	100	100	100				300	
5	300	100	100	100					300
Sujetos	1500	500	500	500	300	300	300	300	300

V1, V2, V3, V4 y V5 indican la versión de examen que contestó cada grupo.

Figura V.2. Estructura de las versiones de examen.

Dificultad	V1	V2	V3	V4	V5	n
BAJA	40	40	40	40	40	200
MEDIA	40	40	40	40	40	200
ALTA	40	40	40	40	40	200
Items	120	120	120	120	120	600

V1, V2, V3, V4 y V5 indican la versión de examen

La siguiente figura indica el contenido homogéneo de las cinco versiones de examen y el orden en que se aplicaron las subpruebas (asignaturas).

Figura V.3. Contenido de las subpruebas.

Subpruebas	Asignaturas	Núm. Items	Numeración
I	Español	24	1 a 24
II	Literatura	12	25 a 36
III	Historia	12	37 a 48
IV	Geografía	12	49 a 60
V	Matemáticas	24	61 a 84
VI	Física	12	85 a 96
VII	Química	12	97 a 108
VIII	Biología	12	109 a 120
Total		120	

Puede observarse que español y matemáticas tuvieron 24 reactivos, lo que constituye el doble del número asignado a las demás asignaturas. La estructura del contenido estuvo dado por el modelo general planteado por el CENEVAL (1994), que es de aplicación común en los exámenes de ingreso al bachillerato en el país.

Las figuras V.4 y V.5 muestran la estructura de las versiones de examen, atendiendo a los tipos de conocimiento y habilidades evaluados.

Figura V.4. Estructura de la prueba por tipo de conocimiento.
Número de reactivos en cada versión.

Categorías	V1	V2	V3	V4	V5
Conceptual	77	79	73	83	79
Factual	22	22	27	21	28
Procedural	21	19	20	16	13
Totales	120	120	120	120	120

Figura V.5. Estructura de la prueba por tipo de habilidad.
Número de reactivos en cada versión.

Categorías	V1	V2	V3	V4	V5
Identificar	58	60	56	60	65
Inferir	29	28	30	29	29
Resolver	33	32	34	31	26
Totales	120	120	120	120	120

Las figuras muestran distribución desigual de reactivos en las versiones. Se advierte que el conocimiento de tipo *conceptual* y la habilidad para *identificar* son notoriamente más extensos, pero mantienen la proporcionalidad en el número de ítems de cada versión.

A.1.2. Resultados totales de los grupos de comparación.

A continuación se presenta una tabla con los indicadores estadísticos de los cinco grupos, señalando la puntuación y la proporción correspondiente al total de aciertos en el total de reactivos.

Tabla V.6. Indicadores estadísticos principales de los cinco grupos.

MUESTRA TOTAL n = 300 N = 1500	PUNTUACIONES					PROPORCIÓN DE ACIERTOS				
	Gpo. 1	Gpo. 2	Gpo. 3	Gpo. 4	Gpo. 5	Gpo. 1	Gpo. 2	Gpo. 3	Gpo. 4	Gpo. 5
Media	60.25	59.72	60.01	60.29	60.46	0.502	0.497	0.500	0.502	0.503
Error típico	1.47	1.45	1.41	1.47	1.52	0.01	0.01	0.01	0.01	0.01
Mediana	54.50	54.00	56.00	53.50	55.50	0.45	0.45	0.47	0.45	0.46
Desviac. Estánd.	25.50	25.13	24.44	25.52	26.36	0.21	0.21	0.20	0.21	0.22
Varianza Mtra.	650.15	631.51	597.45	651.49	694.82	0.05	0.04	0.04	0.05	0.05
Curtosis	-1.17	-1.26	-1.34	-1.29	-1.35	-1.17	-1.26	-1.34	-1.29	-1.35
Rango	94	97	90	102	97	0.78	0.81	0.75	0.85	0.81
Mínimo	22	18	20	18	17	0.18	0.15	0.17	0.15	0.14
Máximo	116	115	110	120	114	0.97	0.96	0.92	1.00	0.95
Suma	18075	17917	18002	18088	18138					

Una primera observación en esta tabla general de resultados es que, aún cuando las medias aritméticas de cada grupo *son desiguales*, no parece haber diferencias importantes entre ellas. El contraste entre la media más alta y la más baja es igual a 0.74 en las puntuaciones y de 0.006 en las proporciones. El único grupo que obtuvo 120 aciertos fue el 4. El menor número de aciertos se observó en el grupo 5, con 17. El grupo 3 tuvo el menor rango; el mayor lo obtuvo el 4. La diferencia entre grupos en el *mínimo de respuestas correctas* fue 5 puntos y 6 en los *máximos*. En ninguno de los dos casos se sobrepasa un 5 % de diferencias, considerando el total de 120 reactivos.

Al tomar como base a las puntuaciones, en los cinco grupos se advierten *curtosis* negativas que oscilaron entre - 1.17 y - 1.35, con una diferencia de .18. El contraste en los valores del *error típico* fue 0.11 y en las *desviaciones estándar* 1.92. Al compararse los datos como proporciones, no se observan diferencias en los valores del *error* y en las *desviaciones* el total es de 0.02. En cuanto a las *medianas*, la diferencia entre mayor y menor es 2.5 puntos y 0.02 en cuanto a la proporción de aciertos.

La siguiente tabla muestra la *frecuencia de respuestas correctas* y la *proporción de aciertos* que obtuvo cada grupo, en las ocho subpruebas aplicadas para evaluar el dominio del contenido.

Tabla V.7. Puntuaciones de los cinco grupos en las ocho subpruebas.

Subpbas	Número y Proporción de Aciertos					Núm. Items
	Gpo. 1	Gpo. 2	Gpo. 3	Gpo. 4	Gpo. 5	
BIO	1851	1985	1872	1960	2019	12
Prop.	0.51	0.55	0.52	0.54	0.56	
ESP	3389	3303	3209	3273	3230	24
Prop.	0.47	0.46	0.45	0.45	0.45	
FIS	1871	1835	1735	1999	1747	12
Prop.	0.52	0.51	0.48	0.56	0.49	
GEO	1825	1860	1975	1886	1903	12
Prop.	0.51	0.52	0.55	0.52	0.53	
HIS	1798	1919	1930	1924	1926	12
Prop.	0.50	0.53	0.54	0.53	0.54	
LIT	1911	1866	1895	1958	1967	12
Prop.	0.53	0.52	0.53	0.54	0.55	
MAT	3458	3325	3393	3195	3444	24
Prop.	0.48	0.46	0.47	0.44	0.48	
QUI	1972	1824	1993	1893	1902	12
Prop.	0.55	0.51	0.55	0.55	0.53	
Total	18075	17917	18002	18088	18138	N = 120
Prop. Med	0.502	0.497	0.500	0.502	0.505	
n	300	300	300	300	300	N = 1500

De acuerdo con la tabla V.2, la diferencia mayor se observó en *física*, con un valor de **0.08**, al comparar las proporciones de aciertos en cada subprueba y la más baja ocurrió en *español*, con valor de **0.02**. Al tomar a la *proporción media de aciertos* de cada asignatura como índice de los niveles de dificultad, pueden contrastarse los resultados de cada versión, procediendo en el orden de menor a mayor dificultad.

Tabla V.8. Orden de las asignaturas en el total de aciertos en cada subprueba

	Versión 1	Versión 2	Versión 3	Versión 4	Versión 5
1	QUÍMICA	BIOLOGÍA	GEO/QUI	FÍSICA	BIOLOGÍA
2	LITERAT.	HISTORIA	HISTORIA	QUÍMICA	LITERAT.
3	FÍSICA	GEO/LIT	LITERAT.	BIO/LIT	HISTORIA
4	GEO/BIO	FIS/QUI	BIOLOGÍA	HISTORIA	GEO/QUI
5	HISTORIA	ESP/MAT	FÍSICA	GEOGRAF.	FÍSICA
6	MATEMÁT.		MATEMÁT.	ESPAÑOL	MATEMÁT.
7	ESPAÑOL		ESPAÑOL	MATEMÁT.	ESPAÑOL
	<i>MEDIA:</i> 0.502	<i>MEDIA:</i> 0.497	<i>MEDIA:</i> 0.500	<i>MEDIA:</i> 0.502	<i>MEDIA:</i> 0.505

Esta tabla indica que las proporciones medias del *total* de aciertos son muy similares en las cinco versiones, pero que los valores asociados a la dificultad de los contenidos en cada versión difieren al interior de cada subprueba, excepto en matemáticas y español, que tuvieron los niveles de dificultad más altos.

A continuación se comparará el total de aciertos obtenido por los sujetos en cada versión de examen, atendiendo al *tipo de conocimiento*. Aunque se indicará la puntuación en cada caso, para el contraste se tomará como base sólo la proporción de aciertos, dado que el número de reactivos de cada tipo es desigual en todas las versiones.

Tabla V.9. Resultados de los cinco grupos en el tipo de conocimiento evaluado.

Total de aciertos: frecuencias y proporciones										
TIPOS DE CONOCIMIENTO										
	V 1	77	V 2	79	V 3	73	V 4	83	V 5	79
CON	FREC	11676	FREC	12040	FREC	11105	FREC	12619	FREC	11952
	PROP	0.505	PROP	0.508	PROP	0.507	PROP	0.506	PROP	0.504
FAC	V 1	22	V 2	22	V 3	27	V 4	21	V 5	28
	FREC	3296	FREC	3259	FREC	4249	FREC	3272	FREC	4325
	PROP	0.499	PROP	0.493	PROP	0.524	PROP	0.519	PROP	0.514
PRO	V 1	21	V 2	19	V 3	20	V 4	16	V 5	13
	FREC	3103	FREC	2618	FREC	2648	FREC	2197	FREC	1861
	PROP	0.492	PROP	0.459	PROP	0.441	PROP	0.457	PROP	0.563

En esta tabla el número en **negritas** indica la cantidad de reactivos en cada grupo, CON= conocimiento conceptual, FAC = factual , PRO = procedural, FREC = frecuencia de aciertos, PROP = proporción de respuestas correctas, y V1 a V5 indican la versión de examen. Los resultados muestran diferencias en los valores de las versiones en los tres tipos de conocimiento. El contraste *más bajo* se advierte en el *conceptual* y el *mayor* en el *procedural*, generado por la versión 5.

Tabla V.10. Ordenación de los tipos de conocimiento: mayor a menor proporción de respuestas correctas en cada versión.

Versión 1	Versión 2	Versión 3	Versión 4	Versión 5
CONCEP.	CONCEP.	FACTUAL	FACTUAL	PROCED.
FACTUAL	FACTUAL	CONCEP.	CONCEP.	FACTUAL
PROCED.	PROCED.	PROCED.	PROCED.	CONCEPT.
MEDIA: 0.502	MEDIA: 0.497	MEDIA: 0.500	MEDIA: 0.502	MEDIA: 0.505

Este orden muestra mayor convergencia en la dificultad de los reactivos que evaluaron el conocimiento *procedural*, ya que el *conceptual* y el *factual* alternaron posiciones en las primeras cuatro versiones. La versión 5 contrasta claramente con las otras cuatro, aunque la diferencia en las medias del total de aciertos no es alta en general.

A continuación se analizarán los resultados observados al contrastar los niveles de dificultad relativos en la medición de los *tipos de habilidad* en las cinco versiones. En la tabla VER1 a VER 5 indican el número de la versión de examen. El número en **negritas** indica la cantidad de reactivos en cada grupo. IDE = habilidad para identificar, INF= inferir y RES para resolver; FREC = frecuencia de respuestas correctas y PROP = proporción de aciertos.

Tabla V.11. Resultados de los cinco grupos en el tipo de habilidad evaluada

TOTAL DE ACIERTOS: Frecuencias y Proporciones										
TIPO DE HABILIDAD										
	VER 1	58	VER 2	60	VER 3	56	VER 4	60	VER 5	65
IDE	FREC	9011	FREC	9492	FREC	8602	FREC	9508	FREC	10041
	PROP	0.517	PROP	0.527	PROP	0.512	PROP	0.528	PROP	0.514
INF	VER 1	29	VER 2	28	VER 3	30	VER 4	29	VER 5	29
	FREC	3792	FREC	3792	FREC	4551	FREC	4146	FREC	4099
	PROP	0.435	PROP	0.451	PROP	0.505	PROP	0.476	PROP	0.471
RES	VER 1	33	VER 2	32	VER 3	34	VER 4	31	VER 5	26
	FREC	5002	FREC	4633	FREC	4849	FREC	4434	FREC	3998
	PROP	0.505	PROP	0.482	PROP	0.475	PROP	0.476	PROP	0.435

Para facilitar el análisis, a continuación se presenta el orden de dificultad que se observó en los tres tipos de habilidad en cada versión, con base en la proporción de aciertos que se obtuvo en cada caso.

Tabla V.12. Ordenación de los tres tipos de habilidad: mayor a menor proporción de respuestas correctas en cada versión.

Versión 1	Versión 2	Versión 3	Versión 4	Versión 5
IDENTIFIC.	IDENTIFIC.	IDENTIFIC.	IDENTIFIC.	IDENTIFIC.
RESOLVER	RESOLVER	RESOLVER	RESOLVER	INFERIR
INFERIR	INFERIR	INFERIR	INFERIR	RESOLVER
<i>MEDIA</i> 0.502	<i>MEDIA</i> 0.497	<i>MEDIA</i> 0.500	<i>MEDIA</i> 0.502	<i>MEDIA</i> 0.505

Puede advertirse que hubo un mayor número de coincidencias en el orden del total de aciertos por *habilidad*, que al comparar el tipo de *conocimiento*. La habilidad que representó menor dificultad fue *identificar* y, excepto en la versión 5, la más difícil fue *inferir*. En general, no es alta la diferencia en las medias.

Las siguientes tablas resumen los valores de las proporciones medias de aciertos de los cinco grupos, para los tipos de conocimiento y de habilidad. En todos los casos se observa que hay una diferencia mínima entre los valores más alto y más bajo de cada versión.

Tabla V.13. Proporción media de aciertos en cada versión. Tipo de conocimiento.

<i>Conceptual</i>	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.51	0.51	0.51	0.51	0.50
<i>Factual</i>	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.50	0.52	0.52	0.52	0.51
<i>Procedural</i>	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.49	0.46	0.44	0.46	0.48

Tabla V.14. Proporción media de aciertos en cada versión. Tipo de habilidad

<i>Identificar</i>	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.52	0.53	0.51	0.53	0.51
<i>Inferir</i>	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.51	0.51	0.51	0.48	0.47
<i>Resolver</i>	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.51	0.48	0.48	0.48	0.51

Para facilitar el análisis, a continuación se presenta el orden de dificultad que se observó en los tres tipos de habilidad en cada versión, con base en la proporción de aciertos que se obtuvo en cada caso.

Tabla V.12. Ordenación de los tres tipos de habilidad: mayor a menor proporción de respuestas correctas en cada versión.

Versión 1	Versión 2	Versión 3	Versión 4	Versión 5
IDENTIFIC.	IDENTIFIC.	IDENTIFIC.	IDENTIFIC.	IDENTIFIC.
RESOLVER	RESOLVER	RESOLVER	RESOLVER	INFERIR
INFERIR	INFERIR	INFERIR	INFERIR	RESOLVER
MEDIA 0.502	MEDIA 0.497	MEDIA 0.500	MEDIA 0.502	MEDIA 0.505

Puede advertirse que hubo un mayor número de coincidencias en el orden del total de aciertos por *habilidad*, que al comparar el tipo de *conocimiento*. La habilidad que representó menor dificultad fue *identificar* y, excepto en la versión 5, la más difícil fue *inferir*. En general, no es alta la diferencia en las medias.

Las siguientes tablas resumen los valores de las proporciones medias de aciertos de los cinco grupos, para los tipos de conocimiento y de habilidad. En todos los casos se observa que hay una diferencia mínima entre los valores más alto y más bajo de cada versión.

Tabla V.13. Proporción media de aciertos en cada versión.
Tipo de conocimiento.

Conceptual	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.51	0.51	0.51	0.51	0.50
Factual	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.50	0.52	0.52	0.52	0.51
Procedural	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.49	0.46	0.44	0.46	0.48

Tabla V.14. Proporción media de aciertos en cada versión.
Tipo de habilidad

Identificar	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.52	0.53	0.51	0.53	0.51
Inferir	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.51	0.51	0.51	0.48	0.47
Resolver	VER 1	VER 2	VER 3	VER 4	VER 5
Media	0.51	0.48	0.48	0.48	0.51

A1.3. Análisis para probar la equivalencia de grupos y versiones de examen.

Una vez que han sido presentados los resultados generales, se procederá a exponer los análisis de varianza de un factor, el *total de aciertos*, expresado como puntuación o proporciones. El objetivo de estas comparaciones es examinar las diferencias entre los grupos y entre las variables consideradas: *contenido, tipo de conocimiento y tipo de habilidad* y determinar su significatividad. El supuesto del que se parte es que todos los sujetos provienen de una misma población, por lo que deberán exhibir patrones de respuesta similares, es decir, no se espera que haya diferencias significativas entre ellos, como efectivamente ocurrió.

Las siguientes tablas indican: fuente de la variabilidad, suma de cuadrados, grados de libertad, promedio de cuadrados, valor de F y su valor crítico.

Tabla V.15. AVAR. Todos los grupos (N=1500). Puntuación total en el examen.

Fuente	Suma Cuadrados	Grados Libert.	Promedio Cuadrados	F	Valor Crítico para F
Entre Gpos	98.75	4	24.69	0.04	2.38
Dentro	964398.98	1495	645.08		
Total	964497.73	1499			

Tabla V.16. AVAR. Todos los grupos (N=1500). Proporción de aciertos.

Fuente	Suma Cuadrados	Grados Libert.	Promedio Cuadrados	F	Valor Crítico para F
Entre Gpos	0.01	4	0.00	0.04	2.38
Dentro	66.97	1495	0.04		
Total	66.98	1499			

Como puede observarse, es indistinto basar el análisis en la puntuación total de aciertos o en su proporción. Para las siguientes tablas se optó por la proporción de aciertos, atendiendo a que los números de reactivos difieren.

Tabla V.17. AVAR. Todos los grupos (N=1500). Habilidad para identificar.

IDENTIF. Fuente	Suma Cuadrados	Grados Libert.	Promedio Cuadrados	F	Valor para F
Entre Gpos	0.065	4	0.016	0.36	2.38
Dentro	67.45	1495	0.045		
Total	67.51	1499			

Tabla V.18. AVAR. Todos los grupos (N=1500). Habilidad para inferir.

INFERIR	Suma	Grados	Promedio		Valor
Fuente	Cuadrados	Libert.	Cuadrados	F	para F
Entre Gpos	0.368	4	0.092	1.65	2.38
Dentro	83.28	1495	0.056		
Total	83.64	1499			

Tabla V.19. AVAR. Todos los grupos (N=1500). Habilidad para resolver.

RESOLVER	Suma	Grados	Promedio			Valor
Fuente	Cuadrados	Libert.	Cuadrados	F	Prob.	para F
Entre Gpos	0.355	4	0.089	1.60	0.17	2.38
Dentro	82.93	1495	0.055			
Total	83.29	1499				

Tabla V.20. AVAR. Todos los grupos (N=1500). Conocimiento factual.

FACTUAL	Suma	Grados	Promedio			Valor
Fuente	Cuadrados	Libert.	Cuadrados	F	Prob.	para F
Entre Gpos	0.108	4	0.027	0.48	0.75	2.38
Dentro	83.29	1495	0.056			
Total	83.39	1499				

Tabla V.21. AVAR. Todos los grupos (N=1500). Conocimiento conceptual.

CONCEPT.	Suma	Grados	Promedio		Valor
Fuente	Cuadrados	Libert.	Cuadrados	F	para F
Entre Gpos	0.003	4	0.00	0.01	2.38
Dentro	67.31	1495	0.045		
Total	67.31	1499			

Tabla V.22. AVAR. Todos los grupos (N=1500). Conocimiento procedural.

PROCEDUR.	Suma	Grados	Promedio		Valor
Fuente	Cuadrados	Libert.	Cuadrados	F	para F
Entre Gpos	0.465	4	0.116	1.92	2.38
Dentro	90.33	1495	0.06		
Total	90.79	1499			

En todos los casos se advierte que F es menor a 2.38, indicando ausencia de diferencias significativas en las comparaciones efectuadas con base en los resultados totales, comprobando la equiparabilidad de los cinco grupos. Se asume que en comparaciones no centradas en los resultados totales podría existir alguna diferencia específica, por lo que el análisis se extendió a otros contrastes.

A.1.4. Sobre la homogeneidad de los grupos: resultados por nivel de desempeño.

A continuación se presentan varias tablas con descriptores estadísticos que atienden al nivel de desempeño (habilidad) de cada uno de los cinco grupos de sujetos, tomando como referencia el número de aciertos y la proporción de respuestas correctas.

Tabla V.23. Total de aciertos en cada grupo: puntuaciones y proporciones.

A) Nivel Alto	PUNTUACIONES					PROPORCIONES				
Gpo. = 100	GPO 1	GPO 2	GPO 3	GPO 4	GPO 5	1	2	3	4	5
Media	91.76	90.67	89.81	91.71	92.47	0.76	0.76	0.75	0.76	0.77
Error típico	0.97	0.84	0.78	0.86	0.84	0.01	0.01	0.01	0.01	0.01
Mediana	88.50	88.50	88.00	89.00	92.00	0.74	0.74	0.73	0.74	0.77
Desviación Estand.	9.72	8.44	7.76	8.64	8.36	0.08	0.07	0.06	0.07	0.07
Varianzade la Mtra.	94.49	71.29	60.26	74.59	69.93	0.01	0.00	0.00	0.01	0.00
Curtosis	-0.05	0.25	-0.09	0.54	-0.50	-0.05	0.25	-0.09	0.54	-0.50
Rango	36	35	30	39	33	0.30	0.29	0.25	0.33	0.28
Mínimo	80	80	80	81	81	0.67	0.67	0.67	0.68	0.68
Máximo	116	115	110	120	114	0.97	0.96	0.92	1.00	0.95
Suma	9176	9067	8981	9171	9247	76.47	75.56	74.84	76.43	77.06

B) Nivel Medio	PUNTUACIONES					PROPORCIONES				
Gpo. = 100	GPO 1	GPO 2	GPO 3	GPO 4	GPO 5	1	2	3	4	5
Media	55.26	55.11	56.59	55.75	57.01	0.46	0.46	0.47	0.46	0.48
Error típico	1.04	1.08	1.06	1.11	1.13	0.01	0.01	0.01	0.01	0.01
Mediana	54.50	54.00	56.00	53.50	55.50	0.45	0.45	0.47	0.45	0.46
Desviación Estand.	10.39	10.77	10.55	11.11	11.29	0.09	0.09	0.09	0.09	0.09
Varianzade la Mtra.	107.93	116.10	111.38	123.52	127.40	0.01	0.01	0.01	0.01	0.01
Curtosis	-0.42	-0.81	-1.08	-1.02	-0.95	-0.42	-0.81	-1.08	-1.02	-0.95
Rango	37	37	39	36	39	0.31	0.31	0.33	0.30	0.33
Mínimo	42	41	40	41	41	0.35	0.34	0.33	0.34	0.34
Máximo	79	78	79	77	80	0.66	0.65	0.66	0.64	0.67
Suma	5526	5511	5659	5575	5701	46.05	45.93	47.16	46.46	47.51

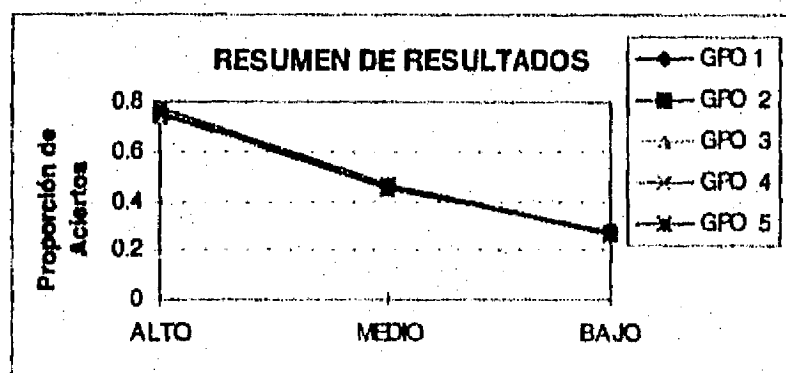
C) Nivel Bajo	PUNTUACIONES					PROPORCIONES					
	Gpo. = 100	GPO 1	GPO 2	GPO 3	GPO 4	GPO 5	1	2	3	4	5
Media		33.73	33.39	33.62	33.42	31.90	0.28	0.28	0.28	0.28	0.27
Error típico		0.48	0.55	0.45	0.47	0.55	0.00	0.00	0.00	0.00	0.00
Mediana		34.00	35.00	34.50	34.00	32.00	0.28	0.29	0.29	0.28	0.27
Desviación Estand.		4.76	5.53	4.53	4.71	5.50	0.04	0.05	0.04	0.04	0.05
Varianza de la Mtra.		22.68	30.56	20.50	22.21	30.25	0.00	0.00	0.00	0.00	0.00
Curtosis		-0.29	-0.16	-0.14	0.01	-0.71	-0.29	-0.16	-0.14	0.01	-0.71
Rango		19	23	20	22	23	0.16	0.19	0.17	0.18	0.19
Mínimo		22	18	20	18	17	0.18	0.15	0.17	0.15	0.14
Máximo		41	41	40	40	40	0.34	0.34	0.33	0.33	0.33
Suma		3373	3339	3362	3342	3190	28.11	27.83	28.02	27.85	26.58

Los datos presentados en la tabla V.23, porciones A, B y C muestran en general dos aspectos que cabe enfatizar. Por una parte, la diferencia *alta* en los totales y en las medias de aciertos de los tres niveles de desempeño. El segundo aspecto es la *homogeneidad* de los cinco grupos en cada nivel de desempeño, siendo más claro al observar los resultados como proporciones. La tabla y gráfica siguientes resumen los resultados.

Tabla V.24. Resumen de resultados en los cinco grupos por niveles de habilidad.

Desempeño	ACIERTOS					ACIERTOS				
	puntuaciones					proporciones				
	GPO 1	GPO 2	GPO 3	GPO 4	GPO 5	GPO 1	GPO 2	GPO 3	GPO 4	GPO 5
ALTO	9176	9067	8981	9171	9247	.76	.75	.74	.76	.77
MEDIO	5526	5511	5659	5575	5701	.46	.45	.47	.46	.47
BAJO	3373	3339	3362	3342	3190	.28	.27	.28	.27	.26
TOTAL	18075	17917	18002	18088	18138	1.50	1.47	1.49	1.49	1.5
Media	60.25	59.72	60.01	60.29	60.46	0.50	0.50	0.50	0.50	0.50
n	300	300	300	300	300	300	300	300	300	300

Gráfica V.1. Resumen de resultados en los tres niveles de habilidad de los cinco grupos.



A continuación se presentan los resultados del desempeño en cada subprueba, considerando los tres niveles de habilidad.

Tabla V.25. Resultados de los cinco grupos, por nivel de desempeño en las subpruebas.

1. ESPAÑOL: habilidad (HAB), suma (SUM)

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	1729	1042	618	3389	1643	1021	639	3303	1585	1024	600	3209
Media	72	43.4	25.8	141	68.5	42.5	26.6	138	66	42.7	25	134
	GRUPO 4				GRUPO 5				AL = Alto			
HAB	AL	ME	BA	TOT	AL	ME	BA	TOT	ME = Medio			
Sum	1713	1022	538	3273	1721	1027	482	3230	BA = Bajo			
Media	13.5	12.9	12.1	34.2	14.6	14.3	10	33.9	TOT = Total			

2. LITERATURA: habilidad (HAB), suma (SUM).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	944	612	355	1911	930	590	346	1866	919	618	358	1895
Media	160	99.7	59.8	319	159	98.5	59.9	317	151	99.1	58.8	308
	GRUPO 4				GRUPO 5				AL = Alto			
HAB	AL	ME	BA	TOT	AL	ME	BA	TOT	ME = Medio			
Sum	954	630	374	1958	948	650	369	1967	BA = Bajo			
Media	165	102	55.9	323	163	102	51.7	316	TOT = Total			

3. HISTORIA: habilidad (HAB), suma (SUM).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	939	548	311	1798	967	601	351	1919	1009	604	317	1930
Media	78.3	45.7	25.9	150	80.6	50.1	29.3	160	84.1	50.3	26.4	161
	GRUPO 4				GRUPO 5				AL = Alto			
HAB	AL	ME	BA	TOT	AL	ME	BA	TOT	ME = Medio			
Sum	968	581	375	1924	997	607	322	1926	BA = Bajo			
Media	80.7	48.4	31.3	160	83.1	50.6	26.8	161	TOT = Total			

4. GEOGRAFÍA: habilidad (HAB), suma (SUM)).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	925	555	345	1825	913	594	353	1860	973	615	387	1975
Media	77.1	46.3	28.8	152	76.1	49.5	29.4	155	81.1	51.3	32.3	165
	GRUPO 4				GRUPO 5				AL = Alto			
HAB	AL	ME	BA	TOT	AL	ME	BA	TOT	ME = Medio			
Sum	932	575	379	1886	939	591	373	1903	BA = Bajo			
Media	77.7	47.9	31.6	157	78.3	49.3	31.1	159	TOT = Total			

5. FÍSICA: habilidad (HAB), suma (SUM).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	954	573	344	1871	930	565	340	1835	881	537	317	1735
Media	79.5	47.8	28.7	156	77.5	47.1	28.3	153	73.4	44.8	26.4	145
	GRUPO 4				GRUPO 5				AL = Alto			
HAB									ME = Medio			
Sum	951	658	390	1999	888	536	323	1747	BA = Bajo			
Media	79.3	54.8	32.5	167	74	44.7	26.9	146	TOT = Total			

6. MATEMÁTICAS: habilidad (HAB), suma (SUM).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	1825	995	638	3458	1774	952	599	3325	1741	1039	613	3393
Media	76	41.5	26.6	144	73.9	39.7	25	139	72.5	43.3	25.5	141
	GRUPO 4				GRUPO 5				AL = Alto			
HAB									ME = Medio			
Sum	1754	921	520	3195	1834	1054	556	3444	BA = Bajo			
Media	73.1	38.4	21.7	133	76.4	43.9	23.2	144	TOT = Total			

7. QUÍMICA: habilidad (HAB), suma (SUM).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	945	592	435	1972	885	570	369	1824	946	624	423	1993
Media	78.8	49.3	36.3	164	73.8	47.5	30.8	152	78.8	52	35.3	166
	GRUPO 4				GRUPO 5				AL = Alto			
HAB									ME = Medio			
Sum	939	566	388	1893	928	579	395	1902	BA = Bajo			
Media	12.8	17.3	12.7	33.3	9.19	16.8	13.3	36.8	TOT = Total			

8. BIOLÓGIA: habilidad (HAB), suma (SUM).

	GRUPO 1				GRUPO 2				GRUPO 3			
	AL	ME	BA	TOT	AL	ME	BA	TOT	AL	ME	BA	TOT
HAB												
Sum	915	608	328	1851	1025	618	342	1985	927	598	347	1872
Media	76.3	50.7	27.3	154	85.4	51.5	28.5	165	77.3	49.8	28.9	156
	GRUPO 4				GRUPO 5				AL = Alto			
HAB									ME = Medio			
Sum	960	622	378	1960	992	657	370	2019	BA = Bajo			
Media	80	51.8	31.5	163	82.7	54.8	30.8	168	TOT = Total			

Tabla V.26. *Conocimiento factual: proporción de aciertos por nivel de desempeño.*

Desempeño	Gpo 1	Gpo 2	Gpo 3	Gpo 4	Gpo 5
ALTO	76.73	77.90	79.04	77.24	78.82
MEDIO	46.45	47.71	50.00	47.48	48.82
BAJO	26.64	29.57	28.33	31.10	26.82
MEDIA	49.94	52.06	52.45	51.94	51.48

Tabla V.27. *Conocimiento conceptual: proporción de aciertos por nivel de desempeño.*

Desempeño	Gpo 1	Gpo 2	Gpo 3	Gpo 4	Gpo 5
ALTO	76.17	76.86	74.88	76.59	76.39
MEDIO	46.91	47.46	48.36	47.30	47.97
BAJO	28.56	28.09	28.89	28.14	26.92
MEDIA	50.61	50.80	50.71	50.67	50.42

Tabla V.28. *Conocimiento procedural: proporción de aciertos por nivel de desempeño.*

Desempeño	Gpo 1	Gpo 2	Gpo 3	Gpo 4	Gpo 5
ALTO	77.29	71.53	69.05	74.50	77.31
MEDIO	42.48	40.00	38.95	40.75	41.85
BAJO	28.00	26.26	24.40	22.06	24.00
MEDIA	49.25	45.93	44.13	45.77	47.72

Tabla V.29. *Habilidad para Identificar: proporción de aciertos por nivel de desempeño.*

Desempeño	Gpo 1	Gpo 2	Gpo 3	Gpo 4	Gpo 5
ALTO	77.90	78.57	75.05	78.55	76.25
MEDIO	48.05	49.42	48.96	49.18	48.85
BAJO	29.41	29.41	29.41	29.41	29.41
MEDIA	51.78	52.46	51.14	52.38	51.50

Tabla V.30. *Habilidad para Inferir: proporción de aciertos por nivel de desempeño.*

Desempeño	1	2	3	4	5
ALTO	77.96	77.96	75.60	73.55	75.28
MEDIO	46.32	46.32	48.17	44.00	44.48
BAJO	29.41	29.41	29.41	29.41	29.41
MEDIA	51.29	51.23	51.06	48.98	49.72

Tabla V.31. *Habilidad para resolver: proporción de aciertos por nivel de desempeño*

Desempeño	1	2	3	4	5
ALTO	78.64	75.13	73.82	75.00	81.08
MEDIO	45.03	43.38	43.29	43.48	47.54
BAJO	29.41	29.41	29.41	29.41	29.41
MEDIA	51.02	49.30	48.84	49.29	52.67

En la siguiente sección se presentarán los resultados de los análisis de varianza hechos para contrastar los resultados de cada nivel de desempeño, aún cuando no se encontraron diferencias significativas entre los sujetos de habilidad alta, media y baja en ninguna de las comparaciones.

A.1.5. Análisis para probar la equivalencia de grupos y versiones por niveles de desempeño.

Para comprobar la equivalencia de las cinco muestras de sujetos en cada rango de desempeño, se hicieron análisis de varianza de un factor con base en el total de aciertos. Se indica el valor de *F* y su nivel crítico.

DESEMPEÑO ALTO

Tabla V.32

	Suma	Grados	Promedio		
Fuente	Cuadrados	Libert.	Cuadrados	F	Valor
Entre Gpos	436.43	4	109.11	1.47	para F
Dentro	36685.24	495	74.11		2.39
Total	37121.67	499	n = 500		

DESEMPEÑO MEDIO

Tabla V.33

	Suma	Grados	Promedio		
Fuente	Cuadrados	Libert.	Cuadrados	F	Valor
Entre Gpos	275.47	4	68.87	0.59	para F
Dentro	58046.96	495	117.27		2.39
Total	58322.43	499	n = 500		

DESEMPEÑO BAJO

Tabla V.34

	Suma	Grados	Promedio		
Fuente	Cuadrados	Libert.	Cuadrados	F	Valor
Entre Gpos	223.11	4	55.78	2.21	para F
Dentro	12494.42	495	25.24		2.39
Total	12717.53	499	n = 500		

En estas tablas se observa que el desempeño *bajo*, que en general mostró patrones de respuesta muy dispersos, estuvo cerca de alcanzar el nivel crítico para *F*, al comparar los 500 sujetos de este nivel.

A continuación se presenta el resultado de los análisis de varianza sobre el *total de aciertos en cada subprueba*, comparando el nivel de desempeño de los cinco grupos. Español y matemáticas tuvieron 24 reactivos, en los demás casos hubieron 12. El número de sujetos en cada grupo es 500.

Tabla V.35. Resultados en español.

GRUPO ALTO					
Origen de variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	647.87	4	161.97	0.47	2.45
Dentro de grupos	39674.46	115	345.00		
Total	40322.32	119			
GRUPO MEDIO					
Origen de variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	12.28	4	3.07	0.01	2.45
Dentro de grupos	34932.92	115	303.76		
Total	34945.20	119			
GRUPO BAJO					
Origen de variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	691.13	4	172.78	0.96	2.45
Dentro de grupos	20599.79	115	179.13		
Total	21290.93	119			

Tabla V.36. Resultados en literatura

GRUPO ALTO					
Origen de variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	67.67	4	16.92	0.13	2.54
Dentro de grupos	7129.58	55	129.63		
Total	7197.25	59			

Tabla V.36 (cont.)

GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	164.00	4	41.00	0.26	2.54
Dentro de grupos	8687.33	55	157.95		
Total	8851.33	59			

GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	41.77	4	10.44	0.15	2.54
Dentro de grupos	3890.17	55	70.73		
Total	3931.93	59			

Tabla V.37. Resultados en historia.

GRUPO ALTO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	253.67	4	63.42	0.48	2.54
Dentro de grupos	7209.67	55	131.08		
Total	7463.33	59			

GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	202.90	4	50.72	0.43	2.54
Dentro de grupos	6534.08	55	118.80		
Total	6736.98	59			

GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	243.73	4	60.93	1.04	2.54
Dentro de grupos	3208.00	55	58.33		
Total	3451.73	59			

Tabla V.38. Resultados en geografía.

GRUPO ALTO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	170.27	4	42.57	0.47	2.54
Dentro de grupos	4959.67	55	90.18		
Total	5129.93	59			
GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	167.67	4	41.92	0.26	2.54
Dentro de grupos	8704.67	55	158.27		
Total	8872.33	59			
GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	104.93	4	26.23	0.22	2.54
Dentro de grupos	6463.25	55	117.51		
Total	6568.18	59			

Tabla V.39. Resultados en física.

GRUPO ALTO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	396.57	4	99.14	0.61	2.54
Dentro de grupos	8913.17	55	162.06		
Total	9309.73	59			
GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	829.23	4	207.31	2.28	2.54
Dentro de grupos	5003.75	55	90.98		
Total	5832.98	59			
GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	274.57	4	68.64	1.05	2.54
Dentro de grupos	3580.17	55	65.09		
Total	3854.73	59			

Tabla V.40. Resultados en matemáticas.

GRUPO ALTO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	292.38	4	73.10	0.49	2.45
Dentro de grupos	17122.42	115	148.89		
Total	17414.80	119			
GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	529.28	4	132.32	0.94	2.45
Dentro de grupos	16159.71	115	140.52		
Total	16688.99	119			
GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	368.95	4	92.24	1.67	2.45
Dentro de grupos	6337.42	115	55.11		
Total	6706.37	119			

Tabla V.41. Resultados en química.

GRUPO ALTO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	215.10	4	53.78	0.24	2.54
Dentro de grupos	12307.08	55	223.77		
Total	12522.18	59			
GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	182.07	4	45.52	0.13	2.54
Dentro de grupos	18795.58	55	341.74		
Total	18977.65	59			
GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	238.67	4	59.67	0.32	2.54
Dentro de grupos	10112.33	55	183.86		
Total	10351.00	59			

Tabla V.42. Resultados en biología.

GRUPO ALTO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	690.90	4	172.73	1.00	2.54
Dentro de grupos	9470.08	55	172.18		
Total	10160.98	59			
GRUPO MEDIO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	166.93	4	41.73	0.28	2.54
Dentro de grupos	8275.25	55	150.46		
Total	8442.18	59			
GRUPO BAJO					
Origen variaciones	Suma cuadrados	GL	Prom.	F	Valor para F
Entre grupos	141.33	4	35.33	0.37	2.54
Dentro de grupos	5199.25	55	94.53		
Total	5340.58	59			

Puede advertirse nuevamente la ausencia de diferencias significativas en las comparaciones realizadas. Al quedar satisfecho el requisito esencial de probar la homogeneidad de las muestras en los aspectos relevantes, se eligió al azar un grupo para los análisis posteriores relacionados en forma directa con el objetivo de este estudio, siendo seleccionado el *grupo 1*. Esta medida se tomó para evitar redundancia innecesaria en el trabajo exhaustivo de identificar los patrones de desempeño inherentes a las respuestas. Las siguientes tablas y gráficas resumen los resultados del grupo 1.

Tabla V.43. Desempeño del grupo 1 en las ocho subpruebas. Proporción de aciertos.

Habil.	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	TOT	n
ALTA	0.72	0.79	0.78	0.77	0.76	0.80	0.79	0.76	0.77	96
MEDIA	0.43	0.50	0.45	0.46	0.41	0.48	0.49	0.50	0.47	107
BAJA	0.26	0.29	0.26	0.29	0.27	0.14	0.36	0.27	0.30	97
Total	0.47	0.53	0.50	0.51	0.48	0.47	0.55	0.51	0.51	300

Cabe señalar que al aplicar puntos de corte para contrastar los niveles de habilidad de los sujetos en los rangos de aciertos *alto* (81 a 120), *medio* (41 a 80) y *bajo* (0 a 40), se producen diferencias pequeñas en el tamaño de cada subgrupo.

Gráfica V.2. Desempeño en las subpruebas por niveles de habilidad.

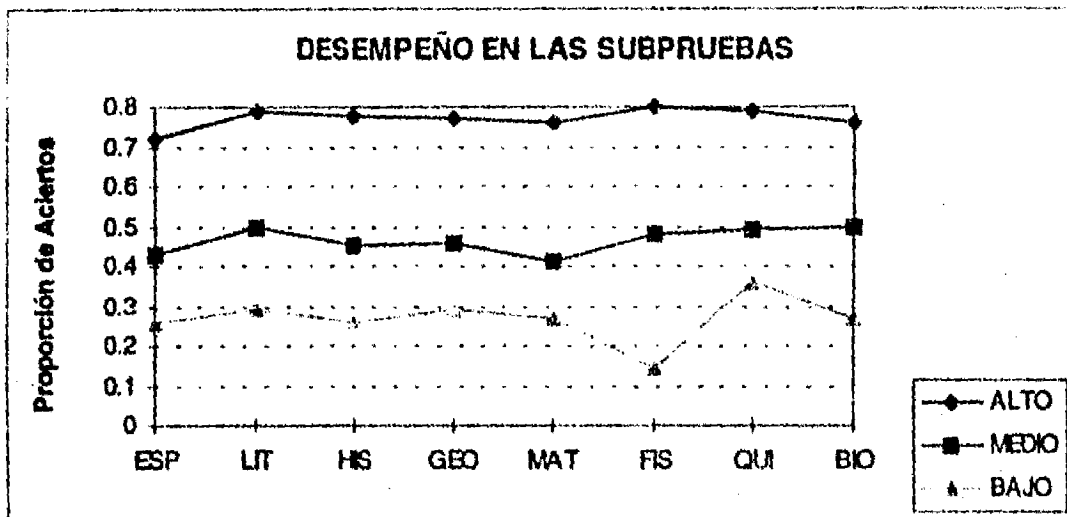
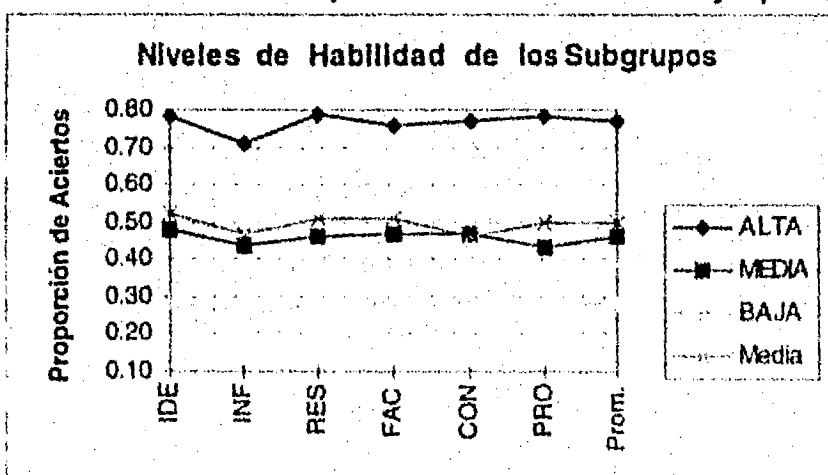


Tabla V.44. Desempeño de los sujetos del grupo 1. habilidades y los tipos de conocimiento

Proporción de aciertos

HABILIDADES			CONOCIMIENTO			TOT	n	Habilidad
IDE	INF	RES	FAC	CON	PRO			
0.78	0.71	0.79	0.76	0.77	0.78	0.77	96	ALTA
0.48	0.44	0.46	0.47	0.47	0.43	0.46	107	MEDIA
0.29	0.25	0.28	0.28	0.15	0.28	0.28	97	BAJA
0.52	0.47	0.51	0.51	0.46	0.50	0.50	300	Total

Gráfica V.3. Desempeño en las habilidades y tipos de conocimiento.



Cabe señalar que al aplicar puntos de corte para contrastar los niveles de habilidad de los sujetos en los rangos de aciertos *alto* (81 a 120), *medio* (41 a 80) y *bajo* (0 a 40), se producen diferencias pequeñas en el tamaño de cada subgrupo.

Gráfica V.2. Desempeño en las subpruebas por niveles de habilidad.

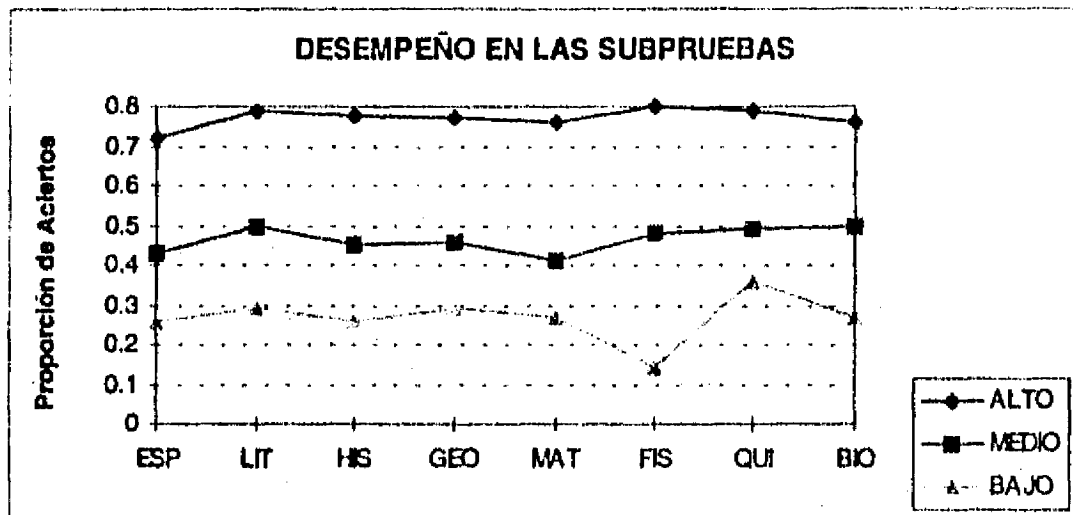
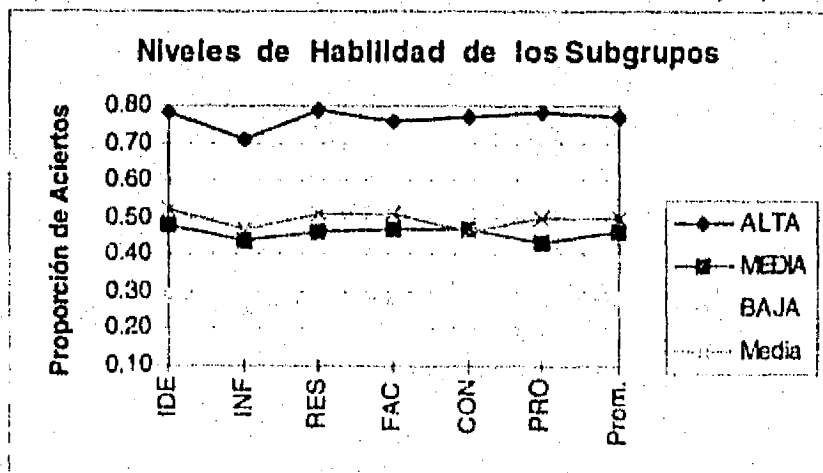


Tabla V.44. Desempeño de los sujetos del grupo 1. habilidades y los tipos de conocimiento

Proporción de aciertos

HABILIDADES		CONOCIMIENTO				TOT	n	Habilidad
IDE	INF	RES	FAC	CON	PRO			
0.78	0.71	0.79	0.76	0.77	0.78	0.77	96	ALTA
0.48	0.44	0.46	0.47	0.47	0.43	0.46	107	MEDIA
0.29	0.25	0.28	0.28	0.15	0.28	0.28	97	BAJA
0.52	0.47	0.51	0.51	0.46	0.50	0.50	300	Total

Gráfica V.3. Desempeño en las habilidades y tipos de conocimiento.



Los resultados de las tablas V.24, V.43 y V.44, así como las gráficas V.1, V.2 y V.3, contienen los elementos básicos para permitir la identificación de los patrones de respuesta, propios de cada nivel explorado. Se puede advertir la presencia de elementos consistentes a través de los indicadores utilizados como definidores de los tres perfiles de desempeño diferencial: alto, medio y bajo, tanto en el manejo del contenido (subpruebas) como en habilidades y tipos de conocimiento.

A2. Análisis psicométrico

Un primer dato a considerar es el sumario estadístico (ITEMAN, 1994), de la respuesta a los 120 ítems que formaron la prueba, por los 300 sujetos del grupo 1. En la tabla del resumen se advierte que la *media de respuestas correctas* está casi diez puntos arriba de 50, lo que habla de un nivel de dificultad medio, moderadamente bajo; la varianza y la desviación estándar indican dispersión alta, en un rango de respuestas correctas de 21 a 116 aciertos.

El valor de la *skinesis* indica que las puntuaciones tendieron a distribuirse simétricamente alrededor de la media, pero al ser positiva señala una abundancia relativa de puntajes en el extremo bajo de la distribución. La *kurtosis* muestra un valor negativo, expresando que la distribución está un tanto aplanada, en relación a la distribución normal, con valor cero.

Tabla V.45. *Resumen estadístico de la escala*

Número de ítems	120
Número de examinados	300
Media	59.930
Varianza	660.992
Desviación estándar	25.710
Skinesis	0.384
Kurtosis	-1.178
Puntuación mínima	21.000
Puntuación máxima	116.000
Mediana	54.000
Coefficiente Alfa	0.965
Proporción media	0.499
Media Item-Tot.	0.435
Media biserial	0.551

El valor *Alfa* de la homogeneidad de la escala indica un coeficiente de confiabilidad .96 alto y muy cercano a 1.0. La *proporción media de respuestas correctas* de todos los ítems de la escala es .49. La *correlación biserial media* entre todos los reactivos de la escala (media ítem-total) es de 0.43 y la *correlación biserial promedio (media biserial)* entre todos los reactivos alcanzó un valor de .55 siendo ambas altas.

El análisis de los ítems mostró que de los 120 reactivos aplicados, 62 de ellos tuvieron una proporción de aciertos entre 0.10 y 0.50, lo que constituyó el 52 % y los 58 reactivos restantes estuvieron entre 0.51 y 0.75.

Los siguientes datos muestran, también en forma sumaria, los valores medios en cada grado de dificultad, la puntuación media, la proporción de aciertos, el índice de discriminación y la correlación punto biserial media de los ítems con el total de aciertos.

Tabla V.46. Resumen de resultados en los grados de dificultad

Grado de Dificultad	Dificultad Media	Puntuación Media	Proporción Media	Discrim. Media	Correl. Media
FÁCIL	56.25	183.93	0.61	0.53	0.45
MEDIO	45.65	151.60	0.51	0.50	0.50
DIFÍCIL	34.03	116.35	0.39	0.47	0.40
Total	45.31	150.63	0.50		

La siguiente tabla indica que más del 90 % de los reactivos tuvo un índice de correlación con el puntaje total, que osciló entre positivo moderado y positivo alto. Dos reactivos tuvieron correlación negativa, indicando un desajuste claro.

Tabla V.47. Rangos de los coeficientes de correlación biserial.

Correlación	de	a	Frec.	%
Negativa			2	1.66
Positiva	0.17	0.30	9	7.5
Positiva	0.31	0.49	75	62.5
Positiva	0.50	0.64	34	28.33
Media		0.52	120	99.99

Al analizar los 10 reactivos que mostraron los niveles de dificultad mayores, se observa que *español* tuvo los índices más altos (seis de diez); en segundo lugar estuvo *biología*, luego *historia*, *química* y *matemáticas*.

Sin embargo, cuando se considera conjuntamente el peso relativo de los reactivos de cada subprueba, el ordenamiento de dificultad se modifica, como se observará más adelante.

De los diez reactivos con dificultad *más alta*, el 24, de español y el 116 de biología, tuvieron valores negativos en la discriminación, y la proporción de aciertos más baja: 0.10 y 0.20, respectivamente. En los ocho ítems, restantes, el nivel de dificultad estuvo entre 0.25 y 0.35, mientras que de los de *menor grado de dificultad*, seis superaron 0.65 en la proporción total de aciertos. En otras palabras, 100 reactivos se distribuyeron sin llegar a valores extremos.

La distribución de aciertos en la prueba se presenta en la tabla V.48, aplicando diez intervalos de 12 puntos, que corresponden en cada caso al 10 % de la prueba. Se observa que 175 sujetos (58 %) se encuentran en la mitad inferior del puntaje máximo, es decir, 60 puntos; y 125 en la mitad superior.

Tabla V.48. Distribución de los sujetos de acuerdo a sus puntuaciones.

Intervalo Aciertos	Frec.	Frec. Acumul.	%	PCT	
1 - 12	0	0	1	0	
13 - 24	7	7	2	2	##
25 - 36	67	74	25	22	#####
37 - 48	62	136	45	21	#####
49 - 60	39	175	58	13	+#####
61 - 72	15	190	63	5	#####
73 - 84	38	228	76	13	#####
85 - 96	45	273	91	15	#####
97 - 108	18	291	97	6	#####
109 - 120	9	300	99	3	+###

|-----+-----+-----+-----+-----+
 5 10 15 20 25
 Porcentaje de Examinandos

El siguiente análisis, basado en RASCAL (1994), permitirá observar los resultados cuando se transforman los puntajes brutos a calificaciones escaladas, con base en unidades *logic*.

Con este fin, se tomará un doble marco de referencia: dificultad de los reactivos y habilidad de los sujetos. Para facilitar el análisis, los reactivos se subdividieron en tres grupos (tablas V.49, V.50 y V.51) asociados al desempeño de los sujetos y la dificultad de los reactivos.

Sin embargo, cuando se considera conjuntamente el peso relativo de los reactivos de cada subprueba, el ordenamiento de dificultad se modifica, como se observará más adelante.

De los diez reactivos con dificultad *más alta*, el 24, de español y el 116 de biología, tuvieron valores negativos en la discriminación, y la proporción de aciertos más baja: 0.10 y 0.20, respectivamente. En los ocho ítems, restantes, el nivel de dificultad estuvo entre 0.25 y 0.35, mientras que de los de *menor grado de dificultad*, seis superaron 0.65 en la proporción total de aciertos. En otras palabras, 100 reactivos se distribuyeron sin llegar a valores extremos.

La distribución de aciertos en la prueba se presenta en la tabla V.48, aplicando diez intervalos de 12 puntos, que corresponden en cada caso al 10 % de la prueba. Se observa que 175 sujetos (58 %) se encuentran en la mitad inferior del puntaje máximo, es decir, 60 puntos; y 125 en la mitad superior.

Tabla V.48. Distribución de los sujetos de acuerdo a sus puntuaciones.

Intervalo Aciertos	Frec.	Frec. Acumul.	%	PCT	
1 - 12	0	0	1	0	
13 - 24	7	7	2	2	##
25 - 36	67	74	25	22	#####
37 - 48	62	136	45	21	#####
49 - 60	39	175	58	13	+#####
61 - 72	15	190	63	5	#####
73 - 84	38	228	76	13	#####
85 - 96	45	273	91	15	#####
97 - 108	18	291	97	6	#####
109 - 120	9	300	99	3	+###

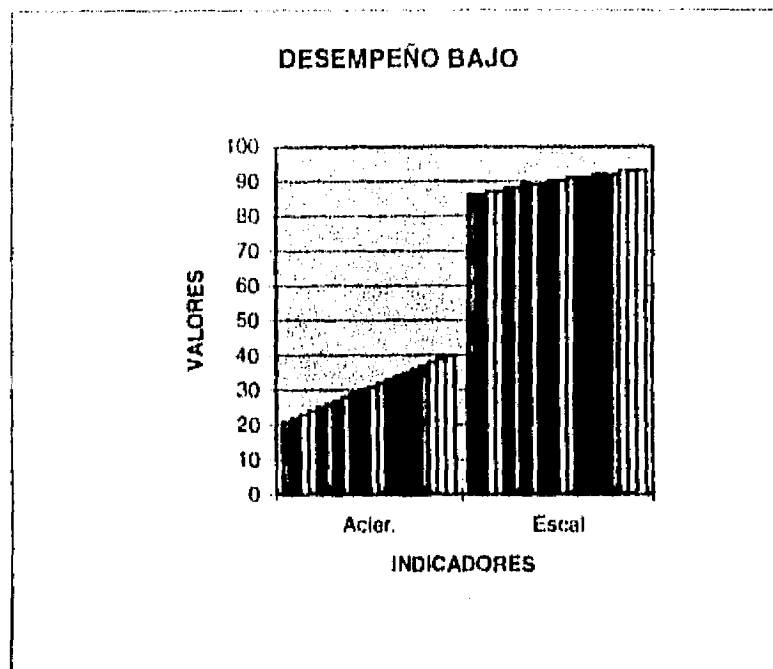
|-----+-----+-----+-----+-----+
5 10 15 20 25
Porcentaje de Examinandos

El siguiente análisis, basado en RASCAL (1994), permitirá observar los resultados cuando se transforman los puntajes brutos a calificaciones escaladas, con base en unidades *logic*.

Con este fin, se tomará un doble marco de referencia: dificultad de los reactivos y habilidad de los sujetos. Para facilitar el análisis, los reactivos se subdividieron en tres grupos (tablas V.49, V.50 y V.51) asociados al desempeño de los sujetos y la dificultad de los reactivos.

Tabla V.49, gráfica V.4. *Indicadores asociados al desempeño bajo.*

Num. Acier.	Punt. Escal.	DESEMPEÑO BAJO			
		Habil.	Frec.	Frec. Acum.	Perc.
21	86	-1.55	1	1	1
22	86	-1.49	3	4	1
23	87	-1.44	3	7	2
24	87	-1.39	0	7	2
25	88	-1.34	4	11	4
26	88	-1.29	2	13	4
27	89	-1.25	2	15	5
28	89	-1.20	4	19	6
29	89	-1.16	1	20	7
30	90	-1.11	5	25	8
31	90	-1.07	8	33	11
32	91	-1.03	6	39	13
33	91	-0.99	10	49	16
34	91	-0.95	6	55	18
35	92	-0.91	5	60	20
36	92	-0.87	14	74	25
37	92	-0.83	5	79	26
38	93	-0.79	7	86	29
39	93	-0.76	8	94	31
40	93	-0.72	6	100	33



La fuente para los indicadores psicométricos son el número de aciertos en la prueba, la puntuación escalada en *logics*, la medida de habilidad, la frecuencia simple y acumulada de los sujetos que obtuvieron cada valor de aciertos y el percentil en que caen los sujetos.

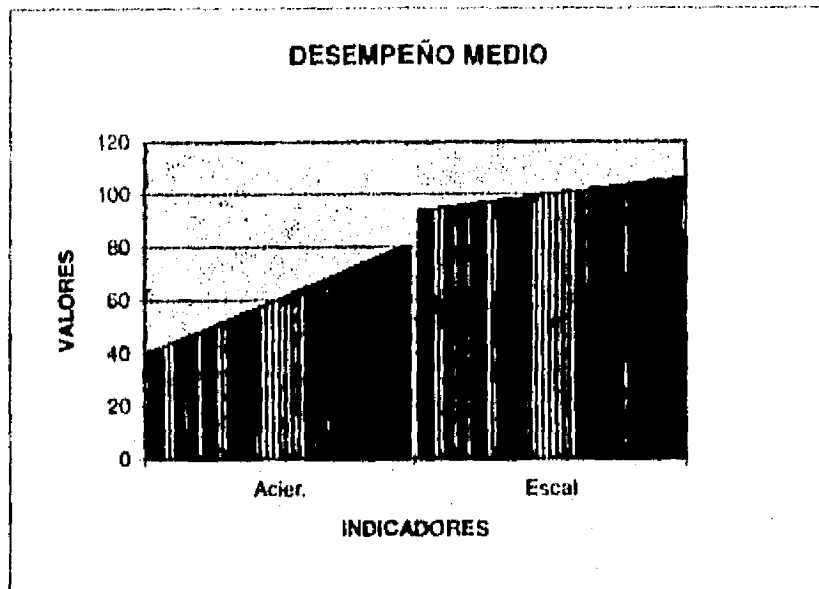
La puntuación escalada se establece por la respuesta de los sujetos a *todos los reactivos*, por lo que no es una suma simple de aciertos, como en la primera columna de las tablas. A partir de este escalamiento se establece el nivel de habilidad, permitiendo ubicar el desempeño de cada sujeto con base en los valores de la tabla.

Un primer aspecto a notar en la tabla V.49 es que los valores asociados a la *habilidad baja* tienen signo negativo, expresando el nivel de *dificultad alta* para los sujetos, en el rango -1.55 a -0.72. El puntaje escalado va de 86 a 93 y los percentiles para el punto de corte van del 1 al 33. Se observa que al aumentar el puntaje crudo hay un ajuste progresivo en el valor del puntaje escalado y del de habilidad. La gráfica 5.3, que acompaña a la tabla, muestra que la tendencia de crecimiento de menor a mayor es común a ambas representaciones.

El segundo conjunto de datos, presentado en la tabla V.50 y gráfica 5.5, los valores de habilidad tienen signo *negativo* hasta 61 aciertos (- 0.01, muy cercano a 0.00), que es justamente el inicio de la mitad superior de la distribución. Se trata de un rango de dificultad-habilidad que comparte características de los dos rangos, el que antecede y el que precede.

Tabla V.50, gráfica V.5 . *Indicadores asociados al desempeño medio.*

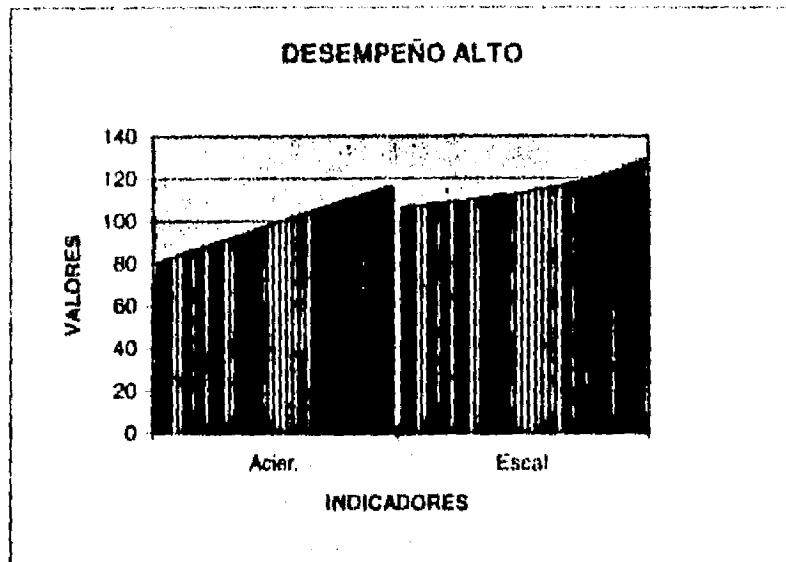
DESEMPEÑO MEDIO					
Num.	Punt.	Habil.	Frec.	Frec.	Perc.
Acier.	Escal.			Acum.	
41	94	-0.68	4	104	35
42	94	-0.65	4	108	36
43	94	-0.61	7	115	38
44	95	-0.58	2	117	39
45	95	-0.54	3	120	40
46	95	-0.51	4	124	41
47	96	-0.47	5	129	43
48	96	-0.44	7	136	45
49	96	-0.41	3	139	46
50	97	-0.37	5	144	48
51	97	-0.34	3	147	49
52	97	-0.31	2	149	50
53	98	-0.27	0	149	50
54	98	-0.24	2	151	50
55	98	-0.21	4	155	52
56	98	-0.17	5	160	53
57	99	-0.14	4	164	55
58	99	-0.11	3	167	56
59	99	-0.07	5	172	57
60	100	-0.04	3	175	58
61	100	-0.01	2	177	59
62	100	0.02	1	178	59
63	101	0.06	0	178	59
64	101	0.09	3	181	60
65	101	0.12	2	183	61
66	101	0.16	2	185	62
67	102	0.19	2	187	62
68	102	0.22	0	187	62
69	102	0.26	0	187	62
70	103	0.29	0	187	62
71	103	0.32	1	188	63
72	103	0.36	2	190	63
73	104	0.39	2	192	64
74	104	0.43	1	193	64
75	104	0.46	2	195	65
76	105	0.50	1	196	65
77	105	0.53	0	196	65
78	105	0.57	1	197	66
79	105	0.60	3	200	67
80	106	0.64	4	204	68



Esto ocasiona una combinación en el paso de *dificultad alta a media* y de *habilidad baja a media* en los sujetos, que sumados a los 100 del bloque anterior totalizan 204. El rango de habilidad se extiende de -0.78 a 0.64. El puntaje escalado tiene un rango de 94 a 106 y los percentiles para el rango de aciertos 41 a 80 van de 35 al 68. Las formas de las gráficas que comparan las puntuaciones son muy semejantes a la del primer bloque, aunque la escala va de 0 a 120.

Tabla V.51, gráfica V.6 . Indicadores asociados al desempeño alto.

DESEMPEÑO ALTO					
Núm.	Punt.		Frec.	Acum.	Perc.
Acier.	Escal	Habil.			
81	106	0.68	4	208	69
82	107	0.72	9	217	72
83	107	0.75	4	221	74
84	107	0.79	7	228	76
85	108	0.83	9	237	79
86	108	0.87	1	238	79
87	108	0.91	5	243	81
88	109	0.95	7	250	83
89	109	1.00	4	254	85
90	109	1.04	2	256	85
91	110	1.08	1	257	86
92	110	1.13	2	259	86
93	111	1.17	7	266	89
94	111	1.22	3	269	90
95	112	1.27	3	272	91
96	112	1.32	1	273	91
97	112	1.37	2	275	92
98	113	1.43	4	279	93
99	113	1.48	1	280	93
100	114	1.54	1	281	94
101	115	1.60	2	283	94
102	115	1.66	2	285	95
103	116	1.73	1	286	95
104	116	1.80	1	287	96
105	117	1.87	1	288	96
106	118	1.94	0	288	96
107	118	2.02	0	288	96
108	119	2.11	3	291	97
109	120	2.20	1	292	97
110	121	2.30	1	293	98
111	122	2.42	1	294	98
112	123	2.54	1	295	98
113	124	2.67	0	295	98
114	126	2.83	2	297	99
115	127	3.01	2	299	99
116	129	3.23	1	300	99



En el *desempeño alto* (tabla V.51 y gráfica V.6), las puntuaciones tienen signo positivo y agrupan 96 sujetos, que sumados a los 204 anteriores totalizan los 300 de la muestra. El rango de habilidad va de 0.68 a 3.23, mostrando valores extremos.

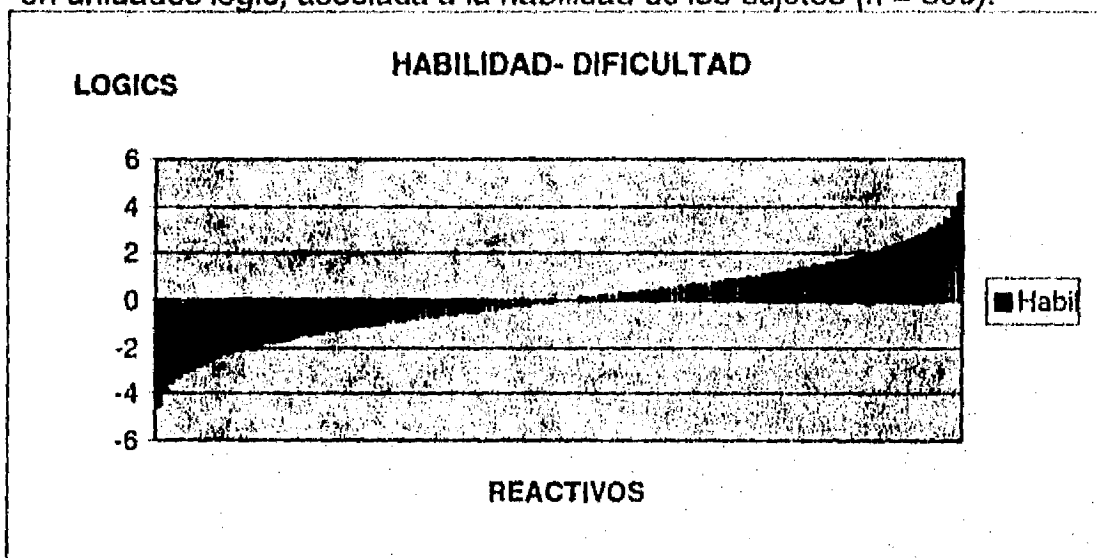
Este aspecto debe señalarse ya que el nivel de habilidad mayor a 2.0 indica un desajuste del instrumento, en este caso a la habilidad de 12 sujetos. Tal indicador funciona de manera análoga al de discriminación que se aplica en los reactivos, pues ambos muestran si los ítems o los sujetos no ajustan con precisión al diseño de la prueba. En sentido inverso, en el primer bloque se ve que 20 sujetos tuvieron menos de 30 aciertos, con calificación

escalada menor a 90, mostrando también otra forma de desajuste: la prueba supera su nivel de habilidad.

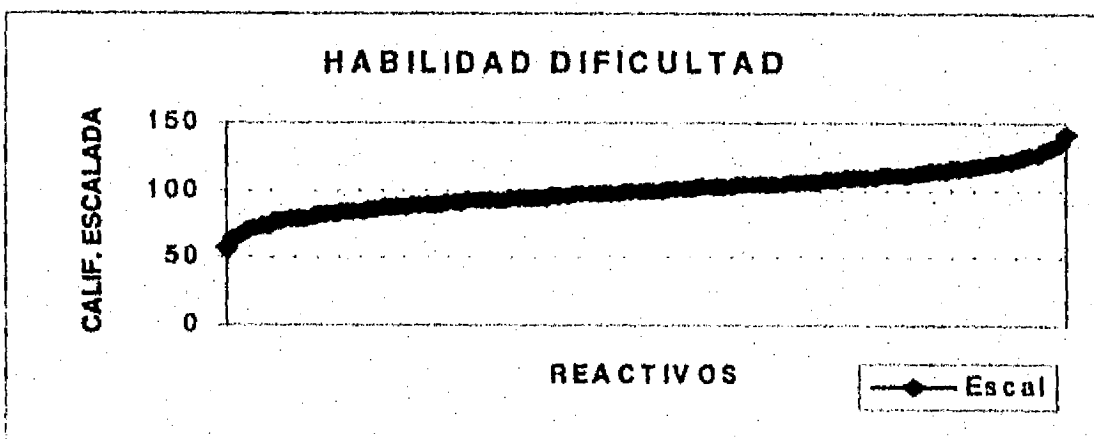
El puntaje escalado en el desempeño alto está en un rango de 106 a 129 y los percentiles para el rango de aciertos 81 (mínimo) a 116 (máximo) va de 69 a 99. En la gráfica 5.5 puede observarse que las puntuaciones crudas mantienen una tendencia lineal creciente y que en los puntajes escalados hay incrementos graduales y menos pronunciados.

Las siguientes gráficas muestran la distribución de la población completa en los indicadores seleccionados: puntuación escalada y nivel de habilidad en la prueba, tomado como base las unidades *logic*.

Gráfica V.7. Distribución de la dificultad de los reactivos medida en unidades *logic*, asociada a la habilidad de los sujetos ($n = 300$).



Gráfica V.8. Distribución de la dificultad escalada de los reactivos asociada a la habilidad de los sujetos.



Con base en estos resultados se concluye que las puntuaciones crudas con rango de 24 a 116 aciertos equivalen en puntuaciones escaladas a un rango de 86 a 129 y que en la habilidad estimada para la muestra fue de -1.55 a 3.23, indicando amplitud y graduación en la dificultad de la

prueba, generando poca redundancia. La representación gráfica de las dos modalidades son congruentes porque las tendencias de las dos distribuciones son similares y muestran relación inversa entre el nivel de habilidad y el de dificultad: cuando la primera aumenta, la segunda disminuye y viceversa.

El último aspecto a considerar en este apartado es la distribución de los 300 sujetos de acuerdo a los puntajes escalados y la distribución por habilidad, estimados a partir de BIGSTEPS (1994), ya que son indicadores de los patrones de desempeño que hasta este momento hemos delineado a partir de la dificultad de los reactivos.

La figura V.6 (página siguiente), muestra cómo se agruparon los sujetos de acuerdo al puntaje total escalado. El sujeto 1 obtuvo el mayor número de aciertos y el 300 el menor. De acuerdo con la distribución de la muestra, los sujetos 1 a 100 representan habilidad alta, del 101 al 200 media y del 201 al 300 baja.

También se observa que los sujetos están agrupados en forma congruente de acuerdo al nivel de habilidad y que no hay una secuencia perfecta del 1 al 300, pero sí hay progresión de mayor a menor puntuación, que está dada no por el puntaje crudo de aciertos, sino por la calificación escalada con base en la respuesta a toda la prueba. Los sujetos con puntajes altos muestran mayor dispersión que los de puntajes medios y bajos, formando estos últimos una agrupación casi rectangular.

Se encuentra también una zona limítrofe en que intersectan el inicio y el final del bloque intermedio con los otros dos, dada la combinación que se da en los grupos medios que comparten algunas características de los sujetos, tanto de habilidad baja como alta. Las zonas sombreadas indican los extremos entre los valores de dificultad en un continuo de -2 a 4, en que la media de la distribución está situada en 0, dado que el rango de puntuaciones va, en orden decreciente, de 3.55 a -1.59.

Identificación de reactivos asociados al nivel de desempeño

El interés central en este trabajo no es explorar hasta el más mínimo detalle las características de la población y de los ítems empleados en la investigación, sino más bien descubrir desde un enfoque psicométrico los patrones inherentes a los niveles de desempeño académico, para someter a prueba la factibilidad de reconocerlos en forma automatizada, con independencia de la muestra estudiada.

Figura V.6. Distribución de la muestra de 300 sujetos por niveles de desempeño.

ITEMS	MAPA DE PERSONAS											
4		001										
		002	003	004								
		005										
3		006										
		007										
		008	010									
		009	011	012								
2		013	015									
		014	016	017	018	020						
		019	021	022	023	024	026	027				
		025	028									
		029	030	031	032	033	034	035	036	037	038	039
		040	041	042	044							
	Q	043	045	046	047	048	049	052	053	091		
1	S	050	051	054	055	056	057	058	059	060	061	062
		063	064	065	066	067	068	069	070	072	074	076
		071	073	075	077	078	079	080	081	082	083	084
		085	086	087	088	089	090	092	093	095	097	098
		094	096	099	100	101	102	103	104	108	112	
	S	105	106	107	109	110	111					
	#####	113										
	#####	114	115	116	117	118	119	120	121	122	132	
0	##### M+M	123	124	125	126	127	128	129	130	131	133	134
		135	136	139								
	#####	137	138	140	141	142	143	144	145	146	147	148
		149	150	151								
	#####	153	154	155	156	157	158	159	160	161	162	163
		164	165	166								
	S	167	168	169	170	171	172	173	174	175	176	177
		178	179	180	181	182	183	184				
	###	185	186	187	188	189	190	191	192	193	194	195
		196	197	198	199	200	202	203	204	205	206	208
		216	217	219								
		201	207	209	210	211	212	213	215	218	220	221
		222	223	224	225	226	227	228	229	230	231	232
		233	234	235	236	237	238	239	240	242	245	
-1	S	241	243	244	246	247	248	249	250	251	252	253
		254	255	256	257	258	259	260	261	262	263	264
		265	266	268	269	270						
	Q	267	271	272	273	274	275	276	277	278	279	280
		281	282	283	287							
		284	285	286	288	289	290	291	295			
		292	292	293	294	296	297	298	299			
		214	300									

-2

Como extensión del estudio se exploró la asociación entre particularidades de los reactivos (contenido, habilidad que miden y tipo de conocimiento), con algunos aspectos del desempeño.

El procedimiento para explorar estas asociaciones estuvo dado por el análisis de regresión por pasos, efectuado en *SPSS for Windows* (1994), atendiendo a la cantidad de varianza explicada (un mínimo de 90%) de los reactivos y tomando como indicadores los contenidos de las subpruebas, habilidad medida y tipo de conocimiento.

Los datos principales se refieren a la población total (n=300), aunque se hará mención de algunos aspectos que se relacionan al desempeño diferencial de los sujetos. Para la revisión de los resultados se integró la información principal en tablas que contienen los datos de mayor interés.

Las primeras tablas están centradas en los contenidos de los reactivos, en relación con el temario en que se basa la estructura de cada subprueba. Se indica la asignatura a que corresponden los reactivos, el número de ítem, la habilidad que miden y el conocimiento que evalúan. En la parte inferior de la tabla se presentan los totales y su distribución.

Por ejemplo, en el caso de *español* los datos muestran que de los 24 reactivos aplicados en la subprueba, 10 explican más del 90 % de la varianza; el orden de ellos en la tabla marca su peso, de mayor a menor. Se advierte que los ítems 10, 3, 21, 4, 5 y 1, es decir, 6 de diez, se relacionan con habilidades verbales (comprensión de lectura y manejo de significado lexical). En lo que se refiere a las habilidades, 7 reactivos se asociaron con los procesos inferenciales y los 10 involucraron el manejo conceptual.

En el caso de *literatura* el tipo de conocimiento dominante también es conceptual, pero la habilidad principal es identificar. Como es de suponerse, El peso mayor está en el manejo de las literaturas clásica, española, etc.

Tabla V.52. Asociaciones de los reactivos con el contenido de las subpruebas: *español*.

ASIG	ITEM	HABILIDAD			CONOCIMIENTO			CONTENIDO	
		1	2	3	1	2	3		
ESP	10	1			1			Semántica	Antónimos, sinónimos
ESP	3	1			1			Lectura de comprensión	Lectura de textos
ESP	13	1			1			Gramática	Enunciados
ESP	12		1		1			Expresión escrita	Acentuación
ESP	21	1			1			Lectura de comprensión	Lectura de textos
ESP	19	1			1			Gramática	Enunc. bímembre
ESP	4	1			1			Lectura de comprensión	Lectura de textos
ESP	15		1		1			Lexicología	Morfemas
ESP	5	1			1			Semántica	Antónimos, sinónimos
ESP	1	1			1			Lectura de comprensión	Lectura de textos
Total	10	1	7	2	10	0	0		

Tabla V.53. Asociaciones de los reactivos con el contenido: *literatura*.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		DE	INF	RES	CON	FAC	PRO	TEMA	SUBTEMA		
LIT	31	1			1				L. Clásicas	Griega y Latina	
LIT	27			1		1			L. Medievales Europeas	Española e Italiana	
LIT	35		1		1				L. Barroca	Barroco literario	
LIT	33	1			1				Generalidades	Formas del lenguaje	
LIT	26			1	1				L. Realista	Caract. Grales.	
LIT	28	1				1			L. Prehispánicas	Náhuatl y Maya	
LIT	25	1			1				Generalidades	Formas del lenguaje	
LIT	36	1				1			L. Española del Renacim.	Caract. Grales.	
Total	8	5	1	2	5	3	0				

Tabla V.54. Asociaciones de los reactivos con el contenido de las subpruebas: *historia*.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		DE	INF	RES	CON	FAC	PRO	TEMA	SUBTEMA		
HIS	47	1			1				El mundo actual	Sist. Socialista	
HIS	39	1				1			México independiente	Conflictos internacionales	
HIS	43	1				1			Porfirismo	Desarrollo Econ. y social	
HIS	46	1				1			El mundo contemporáneo	1a. Guerra Mundial	
HIS	38		1		1				México Prehispánico	Aportes de la Cult. Maya	
HIS	37		1			1			El mundo medieval	Origenes de Europa	
HIS	45		1		1				El mundo moderno	Absolutismo	
HIS	48	1				1			La Conquista de México	Expediciones	
Total	8	5	3	0	3	5					

Tabla V.55. Asociaciones de los reactivos con el contenido: *geografía*.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		DE	INF	RES	CON	FAC	PRO	TEMA	SUBTEMA		
GEO	60	1				1			Situación de México	Límites y Fronteras	
GEO	58		1			1			Zonas económicas	Ubicación en México	
GEO	55	1			1				Población		
GEO	51	1			1				Litosfera	Continentes	
GEO	56	1			1				Medio Ambiente	Recursos Naturales	
GEO	52	1				1			Población		
GEO	49	1			1				La Tierra	Mov. terrestres	
GEO	54	1			1				Medio Ambiente	Hombre-Medio Amb.	
Total	8	7	1	0	5	3	0				

Tabla V.56. Asociaciones de los reactivos con el contenido: matemáticas.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		IDE	INF	RES	CON	FAC	PRO	TEMA	SUBTEMA		
MAT	80			1		1			Álgebra	Monomios, Polinomios	
MAT	68	1						1	Geometría	Polígonos	
MAT	84			1				1	Números Enteros	Representación	
MAT	62	1				1			Números Enteros	Operaciones	
MAT	76			1				1	Números Racionales	Representación	
MAT	67	1						1	Números Naturales		
MAT	78			1				1	Álgebra	Representación	
MAT	77			1				1	Números Racionales	Operaciones	
MAT	66			1				1	Números Enteros	Operaciones	
MAT	65			1			1		Geometría	Paralelismo y Perpendicularidad.	
MAT	63	1					1		Álgebra	Representación	
Total	11	4	0	7	5	0	6				

Tabla V.57. Asociaciones de los reactivos con el contenido: física.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		IDE	INF	RES	CON	FAC	PRO	TEMA	SUBTEMA		
FIS	92			1				1	Mecánica	Mov. Uniforme	
FIS	85			1				1	Mecánica	Fuerza	
FIS	90	1				1			Termodinámica	Calor y Temperatura	
FIS	93			1				1	Termodinámica		
FIS	89			1		1			Mecánica	Mov. Uniforme	
FIS	88		1				1		Ondas	Sonido	
FIS	96			1		1			Termodinámica	Calor y Temperatura	
FIS	95			1				1	Electricidad	Corriente eléctrica	
Total	8	1	1	6	2	2	4				

Tabla V.58. Asociaciones de los reactivos con el contenido: química.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		IDE	INF	RES	CON	FAC	PRO	TEMA	SUBTEMA		
QUÍ	103		1			1			Cálculos químicos	Cálculo de la composición	
QUÍ	106		1				1		Ácidos y Bases	Ácidos y Bases	
QUÍ	101	1				1			Modelos y Caracter. del átomo	Partículas subatómicas	
QUÍ	98	1				1			Modelos y Caracter. del átomo	Partículas subatómicas	
QUÍ	107	1				1			Introducción a la química	Clasificación de sustancias	
QUÍ	97	1				1			Enlaces	Tipos de enlace	
QUÍ	99	1				1			Introducción a la química	Clasificación de sustancias	
QUÍ	105			1				1	Reacciones químicas	Clasificación de reacciones	
Total	8	5	2	1	6	1	1				

Tabla V.59. Asociaciones de los reactivos con el contenido: *biología*.

ASIG	ITEM	HABILIDAD						CONOCIMIENTO		CONTENIDO	
		IDE	INF.	RES	CON	FAC.	PRO	TEMA	SUBTEMA		
BIO	115	1			1				Evolución	Eras geológicas	
BIO	114	1			1				Estudio de la biología	Campo de estudio	
BIO	119	1			1				Niveles de org. biológica	Teoría celular	
BIO	113	1			1				Los seres vivos y la energía	Seres autótrofos y heterótrofos	
BIO	109	1				1			Taxonomía de plantas y animales	Plantas	
BIO	117	1			1				Madurez sexual y desarrollo	Sistema endócrino	
BIO	112	1				1			Niveles de org. biológica		
BIO	120		1		1				Salud	Agentes patógenos	
Total	8	7	1	0	6	2	0				

A continuación se presenta otro conjunto de tablas, centrado en la habilidad que miden los reactivos. En este caso se identifican la habilidad, los ítems con mayor peso en la varianza, las asignaturas con que se relacionan y el tipo de conocimiento involucrado en la asociación. Se revisa primero identificar, luego inferir y por último resolver, tomando como base los 300 sujetos de la muestra.

Cabe señalar que la habilidad para *identificar* está presente en todos los niveles de desempeño y en todas las asignaturas, ya que cognoscitivamente implica, en la mayoría de los casos, las menores demandas. Ciertamente se trata de una habilidad elemental, pero mucho del contenido que se evalúa en las escuelas se queda a este nivel, en tareas que apenas van más allá del mero reconocimiento.

HABILIDAD QUE MIDE EL REACTIVO

Tabla V.60. Asociaciones de los reactivos con la habilidad que miden: *Identificar*.

HABILIDAD	REAC	MATERIA								CONOCIMIENTO			
		ESP	UT	HS	GEO	MT	FIS	DU	BIO	CON	FAC	PRO	
Identificar	115									1		1	
Identificar	114									1		1	
Identificar	48			1								1	
Identificar	83					1					1		
Identificar	47			1							1		
Identificar	30		1								1		
Identificar	43			1								1	
Identificar	54				1						1		
Identificar	51				1						1		
Identificar	109								1			1	
Identificar	29		1									1	
Identificar	14	1									1		
Identificar	118								1		1		
Total	13	1	2	3	2	1	0	0	4	9	4	0	

Tabla V.61. Asociaciones de los reactivos con la habilidad que miden: *inferir*.

HABILIDAD	REAC	MATERIA								CONOCIMIENTO			
		ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	CON	FAC	PRO	
Inferir	9	1									1		
Inferir	3	1									1		
Inferir	57				1						1		
Inferir	35		1								1		
Inferir	7	1									1		
Inferir	21	1									1		
Inferir	38			1							1		
Inferir	10	1									1		
Inferir	58				1							1	
Inferir	19	1									1		
Inferir	88							1				1	
Total	11	6	1	1	2	0	1	0	0	9	2	0	

Tabla V.62. Asociaciones de los reactivos con la habilidad que miden: *resolver*.

HABILIDAD	REAC	MATERIA								CONOCIMIENTO			
		ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	CON	FAC	PRO	
Resolver	92						1						1
Resolver	85						1						1
Resolver	65					1				1			
Resolver	12	1								1			
Resolver	82					1							1
Resolver	69					1							1
Resolver	80					1				1			
Resolver	81					1				1			
Resolver	89						1			1			
Resolver	15	1								1			
Resolver	70					1							1
Resolver	84					1							1
Total	12	2	0	0	0	7	3	0	0	6	0	6	

En los análisis realizados en el presente estudio, pudo observarse que algunos reactivos de *identificar* tuvieron un nivel de dificultad mayor, que otros supuestamente más complejos, por ejemplo, los que se apoyaban en habilidades para *inferir*. Esto ratifica la presencia de interacciones con el tipo específico de contenido y con el conocimiento que se involucra en la evaluación.

A continuación se presentan los resultados de los análisis centrados en los tipos de conocimiento. Se consideró a la muestra completa (n=300). En este caso, cabe señalar que el conocimiento *factual* (hechos, datos, fechas, cantidades, magnitudes), tiene un comportamiento similar a la habilidad de *identificar*, en el sentido de que es común y básico. En contraste, el conocimiento *conceptual* se asocia con la habilidad de *inferir* y el conocimiento *procedural* se relaciona con la habilidad para *resolver*.

Tabla V.63. Asociaciones de los reactivos con el tipo de conocimiento: *factual*.

CONOCIMIENTO	REAC	MATERIA								HABILIDAD			
		ESP.	LIT.	HIS.	GEO.	MAT.	FIS.	QUI.	BIO.	IDE.	INF.	RES.	
Factual	60				1						1		
Factual	109								1		1		
Factual	46			1							1		
Factual	29		1								1		
Factual	39			1							1		
Factual	90						1				1		
Factual	36		1								1		
Factual	48			1							1		
Factual	59				1						1		
Factual	106							1				1	
Factual	37			1								1	
Total	11	0	2	4	2	0	1	1	1	9	2	0	

Tabla V.64. Asociaciones de los reactivos con el tipo de conocimiento: *conceptual*.

CONOCIMIENTO	REAC	MATERIA								HABILIDAD			
		ESP.	LIT.	HIS.	GEO.	MAT.	FIS.	QUI.	BIO.	IDE.	INF.	RES.	
Conceptual	115								1		1		
Conceptual	9	1										1	
Conceptual	12	1											1
Conceptual	80					1							1
Conceptual	3	1										1	
Conceptual	103							1				1	
Conceptual	8	1										1	
Conceptual	31		1								1		
Conceptual	57				1							1	
Conceptual	50				1						1		
Conceptual	96						1						1
Conceptual	47			1							1		
Conceptual	15	1											1
Total	13	5	1	1	2	1	1	1	1	4	5	4	

Tabla V.65. Asociaciones de los reactivos con el tipo de conocimiento: *procedural*.

CONOCIMIENTO	REAC	MATERIA								HABILIDAD			
		ESP.	LIT.	HIS.	GEO.	MAT.	FIS.	QUI.	BIO.	IDE.	INF.	RES.	
Procedural	92						1						1
Procedural	85						1						1
Procedural	84					1							1
Procedural	68					1				1			
Procedural	78					1							1
Procedural	87						1						1
Procedural	82					1							1
Procedural	105							1					1
Procedural	76					1							1
Procedural	77					1							1
Procedural	70					1							1
Total	11	0	0	0	0	7	3	1	0	1	0	10	

Como ha sido planteado en esta primera sección de resultados, el análisis de tipo psicométrico es muy rico en información para formar un esquema completo de los perfiles de desempeño. En el presente estudio, esto se logró al atenderse a los resultados tanto del comportamiento de los reactivos que integraron la prueba, como del de los sujetos evaluados a partir de ella, permitiendo así disponer de un conjunto de puntos de referencia complementarios, que hacen posible enfatizar aspectos que sea de interés particular.

Por ejemplo, permiten describir bajo tres parámetros (manejo de contenido, de habilidades y de tipos de conocimiento) un individuo, un subgrupo (desempeño alto, medio, bajo), un grupo completo e, incluso, a la población de los cinco grupo tomados para la comparación.

Para destacar el perfil diferencial de cada *sujeto*, los resultados obtenidos en este estudio ofrecen una visión integral de su desempeño con base en:

- a) El total de respuestas correctas y la proporción de aciertos en toda la prueba, lo cual permite ubicarlo en un rango de ejecución (alta, media o baja).
- b) El total de respuestas correctas y la proporción de aciertos en cada subprueba, (biología, español, etc.), e incluso en los contenidos del temario relacionados con cada asignatura, atendiendo a la clasificación de cada ítem.
- c) El total de respuestas correctas y la proporción de aciertos en cada tipo de habilidad.
- d) El total de respuestas correctas y la proporción de aciertos en cada tipo de conocimiento.
- e) Su nivel de habilidad global, indicada por el análisis RASCAL.
- f) Su calificación escalada, indicada por el análisis RASCAL.
- g) El percentil en que se ubica, en relación a la muestra de referencia.

Esta disponibilidad de información relevante y útil, fundamenta ratificar la posición adoptada en el presente estudio, desde el planteamiento del problema de investigación: *un perfil completo y realista puede lograrse sólo si se tiene una visión multicomponencial de su estructura.*

Toca ahora revisar las contribuciones que pueden derivarse de la segunda aproximación metodológica adoptada en este estudio, dada por el *procesamiento neurocomputacional*. A manera de presentación general puede señalarse que su manejo no es más árduo o laborioso que el análisis de tipo tradicional. En algunos aspectos es incluso más simplificado y dominarlo como técnica no plantea al investigador demandas de tiempo y esfuerzo inmanejables. Puede decirse que es, en gran parte, un ejercicio interesante, formativo y, sobre todo, muy útil.

B. Resultados del análisis neurocomputacional

Introducción

Puede asumirse que esta sección es la de mayor importancia teórica, metodológica y tecnológica en el presente estudio, ya que se centra en el objetivo de probar la factibilidad de disponer de un reconocedor automatizado de patrones de desempeño académico que satisfaga criterios exigentes de calidad y precisión. Se trata de probar un prototipo cuyas características se asume que son altamente generalizables y su aplicación puede hacerse extensiva al campo del diagnóstico psicológico en sus distintas modalidades.

En este contexto, cabe recordar las razones que harían ventajoso disponer de un reconocedor automatizado de patrones de desempeño. Hay dos aspectos que en especial conviene considerar. El primero se refiere a la precisión del reconocimiento y la calidad del perfil que se obtenga; el segundo tiene que ver con el tiempo, o mejor dicho con la velocidad con que podría efectuarse el análisis. Como podrá observarse en los resultados que se presentan en esta sección, ambos aspectos muestran indicadores satisfactorios.

Los datos que se presentan en esta sección muestran una similitud extrema al comparar los resultados de los modelos psicométrico y neurocomputacional en el reconocimiento de patrones. Los datos de cada caso tienen diferencias mínimas y lo más destacable es que se reproducen los perfiles correspondientes a cada grupo cuando se toma como base *cualquiera* de los tres prototipos de neurocomputador que fueron contruidos para el presente estudio: uno para el grupo 1, otro para el 3 y un tercero para el 5, con el objeto de poder contrastar su funcionamiento.

Como fue expuesto en el capítulo de *método*, el primer neurocomputador fue entrenado con las respuestas de los sujetos del grupo 1, luego se aplicó a reconocer los patrones de los grupos 3 y 5; lo mismo se hizo con el prototipo 3, tomando los grupos 1 y 5 para prueba; y en el grupo 5 acerca de los grupos 1 y 3.

En una corrida alterna y por separado de cada prototipo, aplicada a los sujetos del total de la población abarcada, con $n=1500$ al sumar los cinco grupos, se conservó la precisión y homogeneidad de cada neurocomputador al generar perfiles que permiten identificar a los 500 sujetos de desempeño alto, a los 500 de desempeño medio y a los 500 de desempeño bajo y conocer en cada caso el peso proporcional de cada categoría. Un dato adicional es el de la respuesta en las ocho subpruebas, lo que contribuye a delinear el perfil integral del desempeño de cada sujeto.

El trabajo posterior al procesamiento, consistente en edición de archivos, generar listados y hacer análisis, por ejemplo estadísticos, graficar

y hacer contrastes, se puede realizar con facilidad, ya que se dispone de la información relevante. En sentido estricto, para derivar conclusiones del procesamiento neurocomputacional no se requiere del análisis psicométrico, mientras que, como pudo observarse en este estudio, éste se empobrece sin los datos neurocomputacionales.

La base de la exposición de los resultados estará dada por el desempeño en las ocho subpruebas aplicadas. Por su naturaleza compleja, el *contenido* que es objeto de aprendizaje, es el indicador que se considera de mayor inclusividad y generalidad para evaluar el desempeño en nuestro medio, por ejemplo en exámenes de selección, pues en cierto modo representa una medida de resumen. Este indicador, además de reflejar los aspectos diferenciales del dominio del contenido, involucra también el manejo de habilidades académicas y de tipos de conocimiento asociados al éxito académico, asumiendo que están presentes en los reactivos de la prueba.

Un segundo aspecto a considerar es el señalamiento, expresado también en el diseño del estudio y la porción de método, por utilizar los mismos datos de entrada como fuente de los análisis psicométrico y neurocomputacional, generando en cada caso las matrices que fueran apropiadas al procesamiento.

Este punto llevó a tomar inicialmente los datos de los 300 sujetos de cada grupo y subdividirlos por su desempeño en los tres niveles ya mencionados. Sin embargo, para el procesamiento neurocomputacional se hizo necesario hacer una segunda división al interior de cada grupo, dejando 80 sujetos de cada nivel para la etapa de entrenamiento, con n total de 240 y 20 para la fase de prueba, con $n = 60$.

Este procedimiento se siguió en todos los grupos, pero como ya se había comprobado estadísticamente que entre ellos había alta homogeneidad, para evitar redundancia innecesaria se optó por realizar el análisis completo sólo con los grupos 1, 3 y 5, mismos que servirán como base para el contraste. Esto implica que aún cuando en la comparación de los grupos será desigual el tamaño en las fases psicométrica y neurocomputacional, los aspectos sustanciales serán idénticos.

Otro aspecto relevante se refiere a los números desiguales en los reactivos de las asignaturas español y matemáticas, que tuvieron 24 preguntas, en tanto que las otras seis subpruebas tuvieron 12. Por tal razón, si se toman para el análisis los resultados en la suma de aciertos y en las medias aritméticas, se observará una desproporción, que se desvanece al tomar para la comparación un tercer indicador: la proporción de aciertos.

Para hacer más objetiva la presentación, se ilustrarán todas las fases del procesamiento, iniciando con información de las matrices de entrada de datos y del entrenamiento; se mostrará una porción del listado de *corrida* de

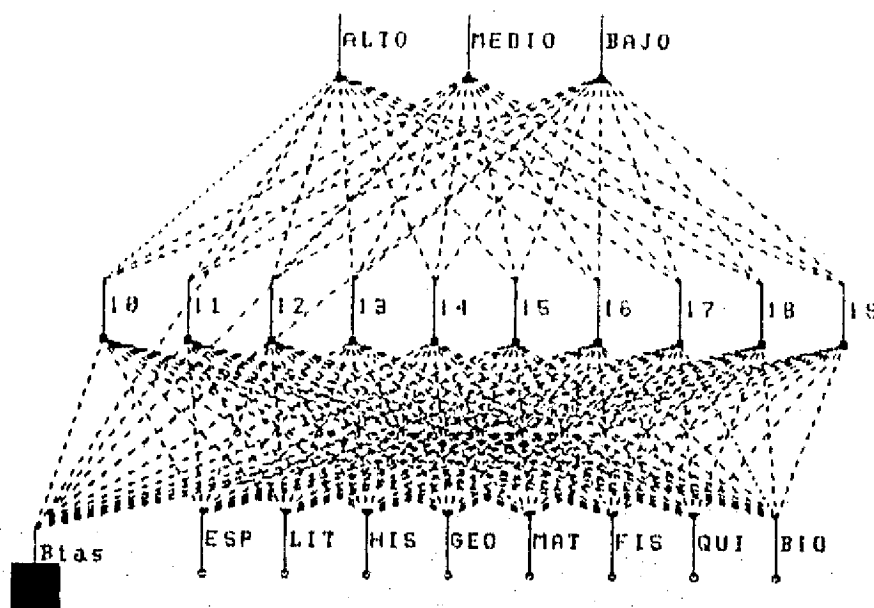
datos y las pantallas de identificación de patrones de desempeño alto, medio y bajo. También se incluirán ejemplos de matrices de salida y prueba, para concluir con la matriz final, que identifica a todos los patrones de interés.

Más adelante se presentarán las tablas de los resultados obtenidos con los dos métodos de análisis para cada grupo, poniendo atención en particular a la proporción de aciertos de cada nivel de desempeño. Para facilitar la comparación se expondrán por separado los datos de los grupos 1, 3 y 5. También se mostrarán los datos que abarcaron a los 1500 sujetos de los cinco grupos.

B.1 Obtención de resultados en el procesamiento neurocomputacional

Figura V.5. Ilustración de la arquitectura de la red neurocomputacional.

PROTOTIPO DE NEUROCOMPUTADOR BACKPROPAGATION-DELTA



Esta figura es igual a la presentada en el segundo capítulo, pero es pertinente incluirla en esta sección porque ayuda a visualizar con claridad todo el proceso. Fue generada en el entorno de *NeuralNetworks* (1989) para construir un prototipo de retropropagador del error aplicando la regla Delta. Muestra tres capas neurodales (input, en la parte inferior de la red, interconexión en la porción intermedia, y output en la porción superior), indicando que el procesamiento va de abajo a arriba (del dato a su integración).

La figura muestra la forma como las capas están interconectadas por las líneas punteadas y puede observarse que: a) hay ocho entradas que se

corresponden con las asignaturas evaluadas; b) en la porción intermedia hay diez elementos de procesamiento con funciones de enlace e integración y c) la salida está formada por tres neurodas, una para cada categoría de estudiante. Hay además un valor constante (*bías*) igual a 1, necesario para normalizar los pesos de la red.

1. Porción de una matriz con datos de entrada

Esta representación mostrará las respuestas registradas en cada reactivo y se expresarán con números: 1 = A, 2=B, C=3, 4=D y 5=E. Los datos de un sujeto a los 120 reactivos están contenidos en dos renglones. El inicio de todo el proceso está dado por esta información, ya que permite saber qué se respondió en cada pregunta, para proceder a identificar los reactivos en que se acertó y, desde luego, en los que se falló. También permite obtener los subtotales en cada subprueba y el total de aciertos en toda la prueba. Este proceso se realiza en forma automatizada, a partir de la lectura óptica de la hoja de respuestas.

Tabla V.61. *Porción de una matriz con datos de entrada.*

544324424143233345314413234422235142314225332124221434222534531214122442
445213523421412525232244245152114512552154515143

544324424143233335314414235422245342314225332124221434222534531214122442
445213523421412525232243245152114512522154535143

531324424143233335314454235422245142314225332124221434242534531214122442
445213523421412325232244245152114512552554555143

544323424143233355314424535122245143314225332123221434222534531213122442
445213523421412525232244245252114512452154515543

534124424143233335314411215422255542314221332124221434222534532214122442
145213523421412525232544245152114512552154545143

2. Porción de una matriz con datos de entrenamiento

Para que el neurocomputador pueda acceder la información es necesario crear un entorno de especificaciones que defina los controles y parámetros necesarios para la operación, como son los valores mínimo y máximo de los datos (en este caso el menor número de aciertos es cero y el mayor es 24) en las columnas que contienen los datos de las subpruebas. Estas definiciones crean el formato de pantalla y, si se desea, puede observarse paso a paso el proceso de entrenamiento, comparando los valores del output que se genera en cada caso y el valor del patrón a igualar.

La estrategia seguida para crear la estructura del análisis es simple: los valores iguales o cercanos a 1 identifican al desempeño alto; los valores iguales o cercanos a 0 identifican al desempeño bajo. Pero, ¿cómo definir el desempeño medio? El valor de .50 sería restrictivo a unos cuantos casos, mientras que los sujetos que son cercanos a .50 alto y a 0.50 bajo son los que en realidad están en el nivel medio. En otras palabras, un sujeto de nivel medio es .50 bajo y .50 alto. Esta estrategia se ajustó perfectamente a la estructura de los datos y favoreció el entrenamiento del neurocomputador.

La importancia de esta matriz es que contiene los parámetros para poder distinguir los perfiles de desempeño, de manera que aún cuando cada conjunto de respuestas difiera en los sujetos, habrá aspectos de comunalidad que permitan identificar y posteriormente reconocer un patrón. A su vez, el neurocomputador registrará los valores específicos de cada lectura, de manera que siempre se podrá individualizar el registro asociado a cada uno de los sujetos.

Figura V.6. Porción de una matriz de entrenamiento

```
input number 1 8
hidden 10
output symbol 1 3
ALTO MEDIO BAJO
display input number 7 1 38 10 2
display output symbol 9 44 3 10 3
display pattern number 7 36 3 10 3
display attribute bold 7 36 3 10 3
display screen 4 24
```

M A T E R I A S

- ESPAÑOL
- LITERATURA
- HISTORIA
- GEOGRAFÍA
- MATEMÁTICAS
- FÍSICA
- QUÍMICA
- BIOLOGÍA

PATRÓN

RESULTADO:

minimum

0 0 0 0 0 0 0 0

BAJO

maximum

24 12 12 12 24 12 12 12

ALTO

facts

20 12 8 8 18 5 8 8

ALTO

14 1 5 0 5 3 4 4

BAJO

20 12 11 10 20 9 10 9

ALTO

14 4 8 7 8 6 2 7

.5 ALTO .5 BAJO MEDIO

9 4 8 8 6 6 5 4

.5 ALTO .5 BAJO MEDIO

ALTO

6 2 1 3 5 7 2 4

BAJO

6 2 3 4 3 2 5 2

BAJO

19 11 9 7 17 11 12 8

ALTO

15 11 5 8 12 10 11 10

ALTO

3. Porción de una matriz con datos de una lectura de datos

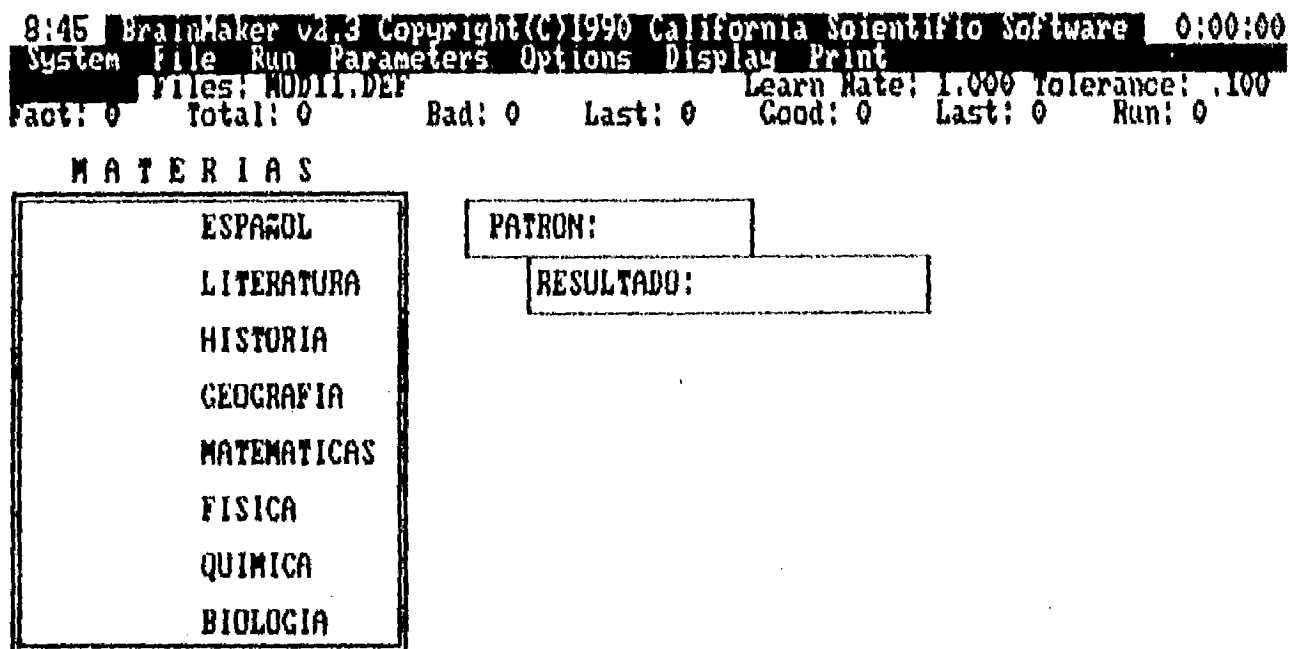
Cuando se hace una corrida de datos es posible crear archivos que muestren los valores generados en cada lectura, indicando el total de hechos leídos (en este caso un hecho equivale a los datos de respuesta de cada sujeto, cuyo total es 240), el total de hechos identificados (*good*), los que no ha reconocido (*bad*), el total no identificado (*badoutputs*), el total acumulativo de hechos no reconocidos (*totalbad*), el margen de tolerancia para el reconocimiento (0.10) y el tiempo de procesamiento en cada corrida.

Puede observarse que en la primera corrida el total de hechos reconocidos es 1; en la tercera lectura aumentó a 15, en la 4 a 33, en la 11 ya había aprendido el patrón de 60 sujetos y en la 15 superó a los 100 sujetos. Se advierte que el tiempo transcurrido para el procesamiento hasta la corrida 15 fue de 31 segundos.

Tabla V.62. Porción de una tabla de lectura de datos.

Run	TotalFacts	Good	Bad	BadOutputs	TotalBad	Tolerance	21:59:08
1	240	1	239	606	606	.100	0:00:01
3	720	15	225	508	1696	.100	0:00:05
4	960	33	207	456	2152	.100	0:00:07
11	2640	60	180	344	4872	.100	0:00:22
15	3600	103	137	307	6153	.100	0:00:31

Figura V.7. Pantalla maestra para la identificación de patronas.



Al observar la figura que muestra la pantalla del neurocomputador, se advierte que en primer término aparecen los datos de los autores del *software* utilizado. En la segunda línea se muestra el menú de opciones para la operación (*system, file, run, parameters, options, display, print*), que son necesarios para cargar archivos, indicar cuál red se desea cargar, seleccionar la forma en que se desea que aparezcan los datos en pantalla, para guardar los datos, etc.

La tercera línea identifica el archivo que se está leyendo, la tasa de aprendizaje aplicada al procesamiento (1.000) y el nivel de tolerancia (0.100). En la cuarta línea se hace un monitoreo de cada corrida: hechos leídos, total leído, hechos no reconocidos (*bad*), el último dato leído (*last*), los datos reconocidos (*good*), el último dato aceptable y el total de corridas.

Es de interés especial la lectura que se hace de cada subprueba, en el recuadro con el nombre de las asignaturas. Y del patrón a igualar en comparación con el *output* que se genera. En las siguientes pantallas se muestra el reconocimiento de patrones con datos reales.

Identificación de un sujeto con desempeño bajo. En este caso pueden leerse los datos de la pantalla que el neurocomputador asocia al patrón de desempeño, dando la lectura en cada subprueba. En español el resultado es 5 aciertos, en literatura es 4, etc. El patrón de entrenamiento muestra un valor de 0.000 y la categoría identificada es "bajo".

Se advierte que se trata del hecho 196 (es decir, del sujeto 196); el total de hechos leídos hasta ese momento fue de 2836, equivalente a 12 corridas (run).

Figura V.8. Pantalla muestra para la identificación de patrones. Perfil bajo.

```

System File Run Parameters Options Display Print
Files: MOD11.DEF Learn Rate: 1.000 Tolerance: .100
Fact: 196 Total: 2836 Bad: 138 Last: 180 Good: 58 Last: 60 Run: 12
  
```

M A T E R I A S

5.000	ESPAÑOL
4.000	LITERATURA
3.001	HISTORIA
4.000	GEOGRAFIA
5.000	MATEMATICAS
1.000	FISICA
6.001	QUIMICA
7.000	BIOLOGIA

PATRON .000

RESULTADO: BAJO

A continuación se presenta una pantalla en que se identifica a un sujeto con *desempeño medio*. Esta pantalla muestra que se trata del sujeto 144 (fact 144). Los valores de las subpruebas son: español 10, literatura 6, historia 6, etc. El patrón a igualar tiene un valor de 0.500 y la categoría asociada es "medio".

Figura V.9. Pantalla muestra para la identificación de patrones. Perfil medio.

```

System File Run Parameters Options Display Print
Files: MOD11.DEF Learn Rate: 1.000 Tolerance: .100
Fact: 144 Total: 3024 Bad: 88 Last: 173 Good: 56 Last: 67 Run: 13
  
```

M A T E R I A S

10.00	ESPAÑOL
6.001	LITERATURA
6.001	HISTORIA
6.001	GEOGRAFIA
6.002	MATEMATICAS
1.000	FISICA
6.001	QUIMICA
6.001	BIOLOGIA

PATRON .500

RESULTADO: MEDIO

Identificación de un sujeto con desempeño alto. Esta última pantalla, corresponde al sujeto 34. Como se observa el recuadro, el patrón tiene un valor de 1.000 y la categoría indicada como resultado es "alto".

Figura V.10. Pantalla muestra para la identificación de patrones. Perfil alto.

```

8:13 PatternMaker V2.3 Copyright(C)1990 California Scientific Software 0:00:16
File Run Parameters Options Display Print
Files: W011.DEF Learn Rate: 1.000 Tolerance: .100
Fact: 34 Total: 754 Bad: 19 Last: 225 Good: 15 Last: 15 Run: 4

```

M A T E R I A S	
15.00	ESPAÑOL
11.00	LITERATURA
10.00	HISTORIA
11.00	GEOGRAFIA
13.00	MATEMATICAS
7.000	FISICA
10.00	QUIMICA
8.002	BIOLOGIA

PATRON 1.000

RESULTADO: ALTO

4. Porción de una matriz de prueba. Cuando en el entrenamiento el total de hechos identificados correctamente es igual al de la muestra de los datos (240), se hacen pruebas para determinar cuál es la precisión del reconocimiento. Para este fin se emplea una parte de la muestra original, que no se haya utilizado en el entrenamiento.

En este caso se trata de 60 sujetos de los 300 originales, ya que el entrenamiento implicó sólo 240. La lectura es similar a la presentada en las pantallas anteriores. Como los datos son conocidos por el investigador y la lectura es extremadamente rápida, casi al momento se puede constatar si el reconocimiento tiene la calidad requerida.

El siguiente paso es presentar a la red un conjunto de datos que corresponden a una muestra que no se empleó para el entrenamiento, por ejemplo los sujetos del grupo 3, asumiendo que el entrenamiento se hizo con los sujetos del grupo 1.

Como se advierte en la siguiente tabla, para la prueba sólo se presentan valores de entrada, tal y como se obtienen de la fuente original, con un mínimo de edición. Aquí puede señalarse la ventaja posterior a todo el trabajo de entrenamiento de la red, ya que para los procesamientos subsecuentes sólo hay que presentar los datos al prototipo y, en un instante, obtener los resultados deseados en la identificación de patrones.

Tabla V.63. *Porción de una matriz de prueba.*

facts run

20	12	8	8	18	5	8	8
17	12	11	9	22	11	11	10
14	1	5	0	5	3	4	4
2	1	6	3	8	4	0	4
5	1	6	4	11	4	6	2
6	1	2	5	4	5	8	7
4	1	1	6	8	5	3	5

5. Porción de una matriz con datos de salida. Los datos que ofrece el prototipo en las primeras ocho columnas son claros en cuanto a los valores procesados. Aparece primero la columna de la primera subprueba (español), luego la segunda (literatura), etc. El orden de los datos permite identificar a qué sujeto se refiere cada renglón.

Después de las subpruebas se presentan los valores asociados al patrón identificado. En el primer caso se indica .000 alto, .000 medio y .999 bajo, expresando que se trata de un sujeto con perfil predominantemente *bajo*. El segundo renglón de refiere a un sujeto con perfil dominante *medio* (.446 alto, .953 medio y .559 bajo). El perfil del tercer sujeto es .988 *alto*, lo que no deja dudas de su ubicación.

Tabla V.64. *Porción de una matriz con datos de salida.*

3.002	2.002	3.001	3.001	11.00	2.002	2.002	.001	.000	ALTO	.000	MEDIO	.999	BAJO
10.00	8.002	4.000	4.000	10.00	4.000	2.002	2.002	.446	ALTO	.953	MEDIO	.559	BAJO
18.00	10.00	9.001	8.002	22.00	11.00	9.001	12.00	.988	ALTO	.001	MEDIO	.011	BAJO
5.000	7.000	8.002	5.002	9.002	6.001	2.002	6.001	.493	ALTO	.985	MEDIO	.499	BAJO
18.00	7.000	11.00	4.000	12.00	5.002	7.000	6.001	.529	ALTO	.945	MEDIO	.481	BAJO
16.00	10.00	10.00	11.00	16.00	9.001	10.00	12.00	.983	ALTO	.002	MEDIO	.015	BAJO
7.004	6.001	4.000	5.002	6.002	3.001	5.002	2.002	.061	ALTO	.030	MEDIO	.939	BAJO
4.004	9.001	6.001	5.002	6.002	3.001	2.002	3.001	.047	ALTO	.022	MEDIO	.951	BAJO
9.002	7.000	5.002	3.001	9.002	5.002	4.000	3.001	.482	ALTO	.961	MEDIO	.517	BAJO
6.002	6.001	2.002	3.001	12.00	4.000	3.001	3.001	.187	ALTO	.344	MEDIO	.817	BAJO
5.000	6.001	2.002	5.002	7.004	3.001	5.002	3.001	.024	ALTO	.004	MEDIO	.975	BAJO
10.00	6.001	3.001	4.000	7.004	8.002	5.002	3.001	.578	ALTO	.992	MEDIO	.424	BAJO

6. Matriz final. El último paso es en realidad de refinamiento de la presentación, puesto que consiste tan sólo en editar los resultados para que su consulta se facilite. Para ello se rotulan las columnas y los valores

asociadas a la categoría pueden redondearse a números enteros, para hacer más discriminativa la clasificación, como 1 y 0.

Tabla V.65. Identificación de patrones: GRUPO 1

ESP	LIT	HIS	GEO	NAT	FIS	QUI	BIO	SUM	PRO	ALT	MED	BAJ	ALT	MEF	BAJ	R
22	10	11	12	20	12	10	11	108	0.90	0.991	0.000	0.009	1	0	0	1
20	10	10	12	24	12	10	10	108	0.90	0.991	0.000	0.009	1	0	0	2
20	8	11	11	23	11	11	10	105	0.88	0.989	0.001	0.011	1	0	0	3
19	11	11	11	21	11	11	9	104	0.87	0.989	0.001	0.010	1	0	0	4
17	12	11	9	22	11	11	10	103	0.86	0.989	0.001	0.010	1	0	0	5
21	11	9	10	20	11	9	11	102	0.85	0.990	0.001	0.010	1	0	0	6
20	11	10	11	20	8	12	10	102	0.85	0.987	0.001	0.012	1	0	0	7
20	12	8	10	21	8	12	10	101	0.84	0.987	0.001	0.013	1	0	0	8
20	12	11	10	20	9	10	9	101	0.84	0.988	0.001	0.011	1	0	0	9
18	8	12	10	20	11	11	10	100	0.83	0.986	0.001	0.013	1	0	0	10
20	11	12	10	16	10	10	10	99	0.83	0.987	0.001	0.012	1	0	0	11
19	8	7	11	22	10	12	9	98	0.82	0.983	0.003	0.016	1	0	0	12
16	11	12	8	20	10	11	10	98	0.82	0.985	0.002	0.014	1	0	0	13
20	11	11	11	17	10	9	9	98	0.82	0.987	0.001	0.011	1	0	0	14
15	11	10	10	18	11	12	11	98	0.82	0.985	0.002	0.014	1	0	0	15
18	8	9	12	23	11	7	9	97	0.81	0.989	0.001	0.010	1	0	0	16
18	10	11	10	19	10	8	11	97	0.81	0.988	0.001	0.011	1	0	0	17
14	10	9	11	20	9	12	11	96	0.80	0.984	0.002	0.015	1	0	0	18

B.2 Resultados del procesamiento neurocomputacional

El análisis se centrará en la comparación de los resultados totales para cada nivel de desempeño en cada una de las subpruebas. El indicador común será la proporción de aciertos obtenida en los perfiles derivados de los dos enfoques: el neurocomputacional y el psicométrico, en los grupos 1, 3 y 5.

Los resultados que obtenidos en el análisis *psicométrico* aportarán el modelo de comparación para juzgar la precisión de los prototipos al igualar el perfil general correspondiente. Para el contraste se presentarán los datos de cada grupo en los tres niveles de desempeño y en cada caso se indicará el número de sujetos, ya sea del grupo o de la muestra evaluada.

En las siguientes tablas aparece en primer término el *modelo a igualar* (*Psicom1*, *Psicom3* y *Psicom5*) en los nueve indicadores que lo componen: asignaturas de cada una de las ocho subpruebas y el total a manera de promedio.

Esta información responde la pregunta: ¿cómo es el perfil de los grupos de desempeño alto, medio y bajo? Se señala después la base de la comparación: *NEUROC 1* identifica el prototipo de neurocomputador 1 y muestra desde esta perspectiva cómo es el perfil del grupo 1 en los niveles de desempeño alto, medio y bajo.

La combinación 3-1 responde a la pregunta: desde la óptica del neurocomputador 3, ¿cómo es el grupo 1? Lo mismo se aplica a la combinación 5-1, ofreciendo la respuesta del prototipo 5 acerca del grupo 1. La última parte de cada segmento de la tabla contrasta los resultados totales de los dos tipos de análisis. Se ofrece también para el modelo en comparación los totales de las combinaciones 3-1 y 5-1. Esta descripción se aplica igualmente a las tablas con los datos de los grupos 3 y 5.

Tabla V.66. Perfil de los sujetos del grupo 1.

A. Sujetos con desempeño alto.

MODELO	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	TOT	N
PSICOM 1	0.72	0.78	0.79	0.77	0.77	0.80	0.79	0.77	0.77	96
Comparación										
NEUROC 1	0.71	0.78	0.78	0.76	0.75	0.79	0.77	0.76	0.76	80
3-1NEU	0.71	0.78	0.78	0.76	0.75	0.79	0.77	0.76	0.76	80
5-1NEU	0.71	0.78	0.78	0.76	0.75	0.79	0.77	0.76	0.76	80

B Sujetos con desempeño medio.

PSICOM 1	0.44	0.52	0.46	0.47	0.42	0.48	0.50	0.51	0.47	103
Comparación										
NEUROC 1	0.43	0.50	0.46	0.46	0.40	0.47	0.48	0.50	0.46	80
3-1NEU	0.43	0.50	0.46	0.46	0.40	0.47	0.48	0.50	0.46	80
5-1NEU	0.43	0.50	0.46	0.46	0.40	0.47	0.48	0.50	0.46	80

C. Sujetos con desempeño bajo.

PSICOM 1	0.27	0.29	0.26	0.29	0.27	0.28	0.36	0.27	0.29	97
Comparación										
NEUROC 1	0.26	0.30	0.26	0.29	0.27	0.29	0.38	0.27	0.29	80
3-1NEU	0.26	0.30	0.26	0.29	0.27	0.29	0.38	0.27	0.29	80
5-1NEU	0.26	0.30	0.26	0.29	0.27	0.29	0.38	0.27	0.29	80

D. Comparación de los totales del grupo

PSICOM 1	0.47	0.53	0.50	0.51	0.49	0.52	0.55	0.52	0.51	300
Comparación										
NEUROC 1	0.46	0.53	0.50	0.50	0.47	0.52	0.54	0.51	0.51	240
3-1NEU	0.46	0.53	0.50	0.50	0.47	0.52	0.54	0.51	0.50	240
5-1NEU	0.46	0.53	0.50	0.50	0.47	0.52	0.54	0.51	0.50	240

Tabla V.67. Perfil de los sujetos del grupo 3.

A. Sujetos con desempeño alto.

MODELO	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	TOT	N
PSICOM 3	0.66	0.77	0.84	0.81	0.73	0.73	0.79	0.77	0.76	100
Comparación										
NEUROC 3	0.65	0.77	0.83	0.82	0.72	0.73	0.78	0.77	0.76	77
1-3NEU	0.65	0.77	0.83	0.82	0.72	0.73	0.78	0.77	0.74	77
5-3NEU	0.65	0.77	0.83	0.82	0.72	0.73	0.78	0.77	0.76	77

B. Sujetos con desempeño medio.

PSICOM 3	0.43	0.53	0.51	0.52	0.44	0.46	0.53	0.51	0.49	98
Comparación										
NEUROC 3	0.43	0.52	0.51	0.52	0.44	0.45	0.53	0.49	0.49	81
1-3NEU	0.43	0.52	0.51	0.52	0.44	0.45	0.53	0.49	0.48	81
5-3NEU	0.43	0.52	0.51	0.52	0.44	0.45	0.53	0.49	0.49	81

C. Sujetos con desempeño bajo.

PSICOM 3	0.25	0.29	0.26	0.32	0.25	0.26	0.34	0.29	0.28	102
Comparación										
NEUROC 3	0.25	0.30	0.26	0.33	0.27	0.26	0.36	0.30	0.29	82
1-3NEU	0.25	0.30	0.26	0.33	0.27	0.26	0.36	0.30	0.28	82
5-3NEU	0.25	0.30	0.26	0.33	0.27	0.26	0.36	0.30	0.29	82

D. Comparación de los totales del grupo

PSICOM 3	0.45	0.53	0.54	0.55	0.47	0.48	0.55	0.52	0.51	300
Comparación										
NEUROC 3	0.44	0.53	0.53	0.55	0.48	0.48	0.56	0.52	0.51	240
1-3NEU	0.44	0.53	0.53	0.55	0.48	0.48	0.56	0.52	0.50	240
5-3NEU	0.44	0.53	0.53	0.55	0.48	0.48	0.56	0.52	0.51	240

Tabla V.68. Perfil de los sujetos del grupo 5.

A. Sujetos con desempeño alto.

MODELO	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	TOT	N	GRUPO
PSICOM 5	0.72	0.79	0.83	0.78	0.76	0.74	0.77	0.83	0.78	100	ALTO
Comparación											
NEUROC 5	0.71	0.79	0.83	0.78	0.76	0.73	0.76	0.82	0.77	80	ALTO
1-5NEU	0.71	0.79	0.83	0.78	0.76	0.73	0.76	0.82	0.76	80	ALTO
3-5NEU	0.71	0.79	0.83	0.78	0.76	0.73	0.76	0.82	0.77	80	ALTO

B. Sujetos con desempeño medio.

PSICOM 5	0.43	0.54	0.51	0.49	0.44	0.45	0.48	0.55	0.49	100	MEDIO
Comparación											
NEUROC 5	0.42	0.53	0.50	0.48	0.44	0.44	0.48	0.54	0.48	80	MEDIO
1-5NEU	0.42	0.53	0.50	0.48	0.44	0.44	0.48	0.54	0.47	80	MEDIO
3-5NEU	0.42	0.53	0.50	0.48	0.44	0.44	0.48	0.54	0.48	80	MEDIO

C. Sujetos con desempeño bajo

PSICOM 5	0.20	0.31	0.27	0.31	0.23	0.27	0.33	0.31	0.28	100	BAJO
Comparación											
NEUROC 5	0.19	0.31	0.27	0.31	0.24	0.28	0.33	0.32	0.28	80	BAJO
1-5NEU	0.19	0.31	0.27	0.31	0.24	0.28	0.33	0.32	0.27	80	BAJO
3-5NEU	0.19	0.31	0.27	0.31	0.24	0.28	0.33	0.32	0.28	80	BAJO

D. Comparación de los totales del grupo

PSICOM 5	0.45	0.55	0.54	0.53	0.48	0.49	0.53	0.56	0.51	300	TOTAL
Comparación											
NEUROC 5	0.44	0.55	0.53	0.52	0.48	0.48	0.52	0.56	0.51	240	TOTAL
1-5NEU	0.44	0.55	0.53	0.52	0.48	0.48	0.52	0.56	0.50	240	TOTAL
3-5NEU	0.44	0.55	0.53	0.52	0.48	0.48	0.52	0.56	0.51	240	TOTAL

Como se muestra en todas las comparaciones, excepto por el número de sujetos, los datos son casi idénticos y sin diferencias significativas. Las implicaciones de estos resultados son muy importantes, ya que partiendo de un prototipo básico puede hacerse la alimentación de datos con poblaciones similares, aplicando un formato simple para que la lectura pueda realizarse.

La confiabilidad del reconocimiento de patrones queda establecida por la convergencia AL 100 % de las lecturas de los tres neurocomputadores.

Otro punto de comparación interesante es tomar como referencia al perfil de los 1500 sujetos evaluados, manteniendo la diferencia de altos, medios y bajos planeada en este estudio para tener $n = 500$ en cada caso.

El contraste consiste en observar la lectura hecha por los prototipos (1,3,5) acerca de la población total. Los indicadores que se aplicarán serán iguales a los de las tablas ya presentadas en esta sección.

Tabla 5.69. Perfil de la población evaluada por cada prototipo

Comparación ESP LIT HIS GEO MAÍ FIS QUI BIO TOT N

A. Sujetos con desempeño alto

1-1500	0.70	0.78	0.81	0.78	0.75	0.77	0.78	0.80	0.77	492
3-1500	0.70	0.78	0.81	0.78	0.75	0.77	0.78	0.80	0.77	492
5-1500	0.70	0.78	0.81	0.78	0.75	0.77	0.78	0.80	0.77	492

B. Sujetos con desempeño medio

1-1500	0.43	0.52	0.49	0.49	0.41	0.48	0.49	0.52	0.48	510
3-1500	0.43	0.52	0.49	0.49	0.41	0.48	0.49	0.52	0.48	510
5-1500	0.43	0.52	0.49	0.49	0.41	0.48	0.49	0.52	0.48	510

C. Sujetos con desempeño bajo.

1-1500	0.24	0.30	0.28	0.31	0.24	0.28	0.34	0.29	0.29	498
3-1500	0.24	0.30	0.28	0.31	0.24	0.28	0.34	0.29	0.29	498
5-1500	0.24	0.30	0.28	0.31	0.24	0.28	0.34	0.29	0.29	498

D. Comparación de los totales de la población completa.

1-1500	0.46	0.53	0.53	0.52	0.47	0.51	0.53	0.54	0.51	1500
3-1500	0.46	0.53	0.53	0.52	0.47	0.51	0.53	0.54	0.51	1500
5-1500	0.46	0.53	0.53	0.52	0.47	0.51	0.53	0.54	0.51	1500

Se observa nuevamente que no hay diferencias en los perfiles, lo que significa que cualquiera de los tres prototipos, entrenado con 240 sujetos, tiene una generalidad probada para identificar los patrones de poblaciones mucho mayores, en este caso compuesta por 1500 sujetos.

Un aspecto interesante es que los tres neurocomputadores fueron coincidentes al identificar 492 sujetos de nivel alto, 510 de nivel medio y 498 del bajo.

Esa lectura de los datos fue correcta y exacta, ya que aún cuando de inicio se planeó que cada grupo de 300 sujetos debería tener 100 de los tres niveles de desempeño, no se cumplió del todo al hacer la selección. Por esta razón los tamaños de las muestras difirieron en proporciones mínimas, pero conservaron su equivalencia en las variables relevantes, como fue comprobado en las análisis que se efectuaron con esa finalidad.

La precisión observada indica que, lo que realmente cuenta para el procesamiento son los datos, demostrando así la objetividad del sistema.

CAPITULO VI

Discusión

A continuación se tratarán puntos centrales del estudio realizado, en términos de la discusión de los resultados obtenidos a partir de la aplicación de los indicadores manejados para identificar los patrones del desempeño, atendiendo al análisis psicométrico en primer lugar, para abordar después los aspectos pertinentes al procesamiento neurocomputacional. Por la naturaleza y objetivos de esta investigación, es especialmente relevante resaltar el comportamiento de los sujetos por niveles de habilidad, más que el de la población; y el de los reactivos por sus grados de dificultad, más que el de la prueba en su conjunto.

A. DISCUSIÓN DE RESULTADOS EN LA PORCIÓN PSICOMÉTRICA.

El desarrollo de esta sección considerará los siguientes dos aspectos, en congruencia con la estructura seguida en los análisis:

1. Discusión de los resultados globales, basada en los perfiles de desempeño asociados a la habilidad de los sujetos y la dificultad de los reactivos.
2. Discusión de los perfiles de desempeño asociados al contenido de la evaluación, tipos de habilidad y tipos de conocimiento.

A.1. Discusión de los resultados globales.

La evaluación planeada y ejecutada con base en el modelo descrito en el capítulo I, incorporó los componentes de *una estrategia metodológica multicomponential y multidimensional*. Este modelo se aplicó a la evaluación integral del desempeño escolar, orientada a identificar perfiles diferenciales asociados a las respuestas, en un examen estándar que midió: a) manejo de contenidos temáticos de ocho asignaturas de los programas de estudio del nivel de enseñanza medio superior, b) tres habilidades cognoscitivas y c) tres tipos de conocimiento.

La estrategia metodológica adoptada en el estudio para establecer los perfiles diferenciales demarcó rangos de desempeño dependientes del total de aciertos en una prueba con 120 reactivos. El nivel *bajo* quedó ubicado de 0 a 40 aciertos, el *medio* de 41 a 80 y el *alto* de 81 a 120.

De igual forma, la especificación de la dificultad de los reactivos, asociada a la habilidad de los sujetos, fue definida de manera empírica y en forma previa a la aplicación, para distinguir con objetividad tres grados: *fácil*,

medio y difícil, determinando que su ordenamiento dependería del número total de aciertos, en tres porciones con 40 reactivos.

Con el objetivo de disponer de marcos de comparación apropiados para el análisis de los patrones, se utilizaron cinco grupos independientes ($N=1500$, $n= 300$), distribuyendo a los sujetos de acuerdo a su número de aciertos total en la prueba, en tres subgrupos: 100 de habilidad alta, 100 media y 100 baja. A estos grupos se les aplicaron versiones de examen similares, en la medida en que compartían 77 reactivos comunes. La siguiente tabla resume los resultados generales de los cinco grupos.

Tabla VI.1. Resumen de resultados en los cinco grupos por niveles de habilidad.

MEDIA DE ACIERTOS					Nivel de Desempeño
GPO 1	GPO 2	GPO 3	GPO 4	GPO 5	
76.47	75.56	74.84	76.43	77.06	ALTO
46.05	45.93	47.16	46.46	47.51	MEDIO
28.11	27.83	28.02	27.85	26.58	BAJO
50.21	49.77	50.00	50.24	50.38	Media
300	300	300	300	300	n

La primera parte del análisis tuvo como meta probar equivalencia de sujetos e instrumentos. Dicha meta se alcanzó satisfactoriamente, en la medida en que fue avalada con suficiencia por los análisis de varianza, tanto por grupo ($n = 300$, $N=1500$) como por niveles de desempeño (entre los sujetos *altos*, $n = 500$; *medios*, $n = 500$ y *bajos*, $n = 500$), como fue mostrado en el capítulo V, porción de resultados del análisis psicométrico.

Cabe asumir que si se hubieran encontrado diferencias significativas, lo cual no ocurrió, estarían presentes problemas serios en la equiparabilidad de los grupos, ya que estaría abierta una fuente importante de invalidez.

Estos datos fundamentaron que, para los análisis subsecuentes, se utilizara sólo uno de los cinco grupos, con el propósito de evitar redundancia innecesaria al efectuar una exploración más detallada de los patrones de interés. Al azar resultó seleccionado el grupo 1.

La media de aciertos de los 300 *sujetos* del grupo 1 fue 60.25, la mediana 54.50, la moda 85, la desviación estándar tuvo un valor de 25.50 y en la varianza de la muestra fue 650.15. El mínimo de aciertos fue 22 y el máximo 116. En relación a los indicadores estadísticos de la *prueba*, la media de aciertos en los 120 reactivos fue 150.63, la media tuvo un valor de 151.50 y la moda fue 156. La desviación estándar fue 32.47 y la varianza 1054.40. El mínimo de aciertos fue 30 y el máximo 228. El total de aciertos fue 18,075.

La tabla VI.2 resume los resultados globales del grupo 1, basando la comparación en la *dificultad* de los reactivos y la *habilidad* de los sujetos. En ella se muestran datos que en la evaluación hicieron evidentes los perfiles diferenciales, ya que conjuga los dos parámetros aplicados, expresados en la suma del total de aciertos, el porcentaje de aciertos y la dificultad media de los reactivos.

Puede observarse en la tabla que el número de aciertos disminuyó consistentemente conforme aumentó la dificultad de la prueba y, desde luego, en la medida en que también decreció la habilidad de los sujetos. El índice de dificultad de los reactivos, calculado a partir del total de aciertos obtenidos por los sujetos en la prueba, indica que el valor más bajo se da en los reactivos que son más difíciles.

En este sentido, es común en estudios como el realizado, que se busque situar la dificultad de los reactivos en valores alrededor de 50, con 15 puntos hacia arriba y 15 puntos hacia abajo, para evitar que se incluyan reactivos con valores extremos (muy fáciles o muy difíciles), porque de acuerdo con esta concepción, se generaría una pérdida del poder de discriminación. En los análisis tipo Rasch esta dificultad se ubica en el punto 0 en la una escala que, dependiendo de la amplitud de la dispersión, suele estar entre -3 y + 3. Cuando se le compara con una escala de dificultad formada con los puntajes crudos, se observa que puede admitir reactivos con valores tan extremos como 10 y 80, siempre y cuando estén ajustados a los parámetros de la habilidad de los sujetos, en una correlación ítem-prueba.

Sin embargo, cuando se aplica el criterio de límites de dificultad entre 35 y 65, los mejores reactivos calculados con la técnica de Rasch quedan comprendidos en esa distribución y sus índices de discriminación son muy apropiados.

Tabla VI.2. *Relación entre los tres niveles de habilidad y los tres grados de dificultad. Número de aciertos.*

Dificultad	Habilidad			Total	%	Dific. Media	n Suj.	n Items
	ALTA	MEDIA	BAJA	Aciert.				
FÁCIL	3441	2412	1504	7357	40.70	56.25	96	40
MEDIO	3170	1837	1057	6064	33.55	45.65	107	40
DIFÍCIL	2565	1276	813	4654	25.75	34.03	97	40
Total	9176	5525	3374	18075			300	120
%	50.77	30.57	18.67	$\chi^2 = 91.245 (4 \text{ gl}), p < 0.001, \text{bilateral}$				

La χ^2 (chi cuadrada) calculada en un arreglo de 3x3, contrastando a la vez los niveles de habilidad y los grados de dificultad, produjo un valor de

probabilidad asociado a diferencias significativas muy altas, confirmando que las respuestas de los sujetos obedecen a patrones bien definidos.

Para analizar con mayor precisión los patrones de respuesta de los sujetos y el comportamiento de los reactivos de la prueba, se procedió a construir la matriz que se reproduce en la figura VI.1.

La fuente de este paradigma está dada por los análisis que se apegan a los modelos IRT y que toman las bases conceptuales de la psicometría derivada de Rasch. Dichos modelos plantean una exigencia metodológica, que fue tomada como eje central en este trabajo, al señalar que *debe existir correspondencia directa entre el nivel de habilidad que los sujetos exhiben en una prueba y el grado de dificultad de los reactivos*.

El ordenamiento de los sujetos en este estudio siguió la estructura de la matriz planteada, atendiendo al número de aciertos, como primer paso para analizar las trayectorias de respuesta indicadoras de los perfiles de desempeño.

Esta estrategia aportó una herramienta poderosa para analizar los ajustes y desajustes de la prueba en dos sentidos: la dificultad de los ítems a los niveles de habilidad de los sujetos y de éstos al modelo de la Teoría de Respuesta de los Reactivos (IRT), que asume que la habilidad más alta se asocia consistentemente a los aciertos en los ítems de mayor dificultad.

Figura VI.1. Paradigma sujeto-reactivo

SUJ.	REACTIVOS									Aciertos	NIVELES DE HABILID.
	1	2	3	4	5	6	7	8	9		
1	1	0	0	0	0	0	0	0	0	1	BAJO
2	1	1	0	0	0	0	0	0	0	2	
3	1	1	1	0	0	0	0	0	0	3	
4	1	1	1	1	0	0	0	0	0	4	MEDIO
5	1	1	1	1	1	0	0	0	0	5	
6	1	1	1	1	1	1	0	0	0	6	
7	1	1	1	1	1	1	1	0	0	7	ALTO
8	1	1	1	1	1	1	1	1	0	8	
9	1	1	1	1	1	1	1	1	1	9	
Aciertos	9	8	7	6	5	4	3	2	1	TOTAL	
	BAJO			MEDIO			ALTO				
	NIVELES DE DIFICULTAD										

1 = Acierto
0 = Error

La utilidad de esta matriz se ilustra en la siguiente figura.

Figura VI.2. Paradigma sujeto-reactivo

		REACTIVOS									
SW.	1	2	3	4	5	6	7	8	9	Aciertos	
1	1	1	1	1	1	0	0	0	0	5	
2	1	1	1	0	1	1	0	0	0	5	
3	0	1	0	1	0	1	0	1	1	5	
4	0	0	0	1	1	1	1	1	0	5	
5	0	0	1	1	0	1	0	1	1	5	
		BAJO			MEDIO			ALTO			
		NIVELES DE DIFICULTAD									

La figura contiene los patrones de respuesta de cinco sujetos en nueve ítems, indicando que todos ellos obtuvieron un total de cinco aciertos. Se asume que los reactivos están ordenados de menor a mayor dificultad, de acuerdo a los resultados de la población completa que fue evaluada. Si, como es común, se prestara atención sólo al número total de aciertos, estos datos indicarían que los cinco sujetos no difirieron en habilidad. Pero ¿sería válida esa conclusión?

El sujeto 1 fue el más congruente de este grupo: acertó en los cinco reactivos más fáciles y falló en los cuatro de mayor dificultad. El sujeto 2 es un tanto parecido, pero falló en el reactivo 4 (más fácil que el 6) en que cabría esperar que acertara y acertó en uno (el ítem 6, más difícil que el 4), en que podría esperarse que fallara. El sujeto 3 se muestra errático, pues en un reactivo falla y al siguiente acierta y además responde correctamente el reactivo más difícil (9), a la vez que no acierta en el más fácil (1). El sujeto 4 presenta un perfil que se parece al de los alumnos que copian cuando tienen oportunidad: falla en los tres primeros reactivos más fáciles y en los siguientes acierta. Por último, el sujeto 5 da la impresión de que contestó al azar.

Puede advertirse que las conclusiones acerca del nivel de habilidad de los sujetos, cambian a partir del análisis de las trayectorias de respuesta. De alguna forma se está indicando que los sujetos son inconsistentes. Pero ¿es correcta esta nueva suposición? La respuesta dependería de los que se supiera acerca de las características del instrumento aplicado. Si se tratara de un instrumento cuyos reactivos no fueron calibrados, lo que en realidad se estaría mostrando son los desajustes propios de una prueba defectuosa.

Si se quisiera reproducir en este punto la matriz con las respuestas de los 300 sujetos del grupo 1 en los 120 reactivos aplicados, con el objeto de

analizar la existencia de incongruencias, como las que se ilustraron con los datos ficticios mostrados en la figura, la tarea de revisión sería interminable.

Para este propósito sirven los análisis de Rasch, ya que se pueden hacer automatizadamente (por ejemplo: BIGSTEPS, 1994; RASCAL, 1992). Los valores que se calculan señalan desajustes de los *reactivos*, a partir de las trayectorias de respuesta de todos los sujetos; y de los *sujetos*, a partir de los valores de los reactivos, calculados con base en los aciertos y errores en las respuestas. Cabe tener presente que los análisis de tres parámetros indican la dificultad del ítem, el poder de discriminación y la probabilidad de respuestas generadas al azar. La habilidad de los sujetos se expresa en una calificación escalada en unidades lógicas y la dificultad de los reactivos en un rango que, por lo general, va de menos tres a más tres.

En este estudio, los resultados mostraron que 118 de los 120 ítems separaron con precisión a los sujetos de habilidad alta de los de habilidad baja. También se observó que 12 reactivos estuvieron por debajo de la habilidad de los sujetos (muy fáciles) y 20 por encima de su habilidad, es decir, dificultad muy alta. En sentido estricto, puede asumirse que 88 de los 120 reactivos (73 %) tuvieron un comportamiento óptimo, considerando el conjunto de las cinco versiones aplicadas.

Este resultado es congruente, si se tiene en cuenta que 77 reactivos fueron comunes a los cinco grupos, generando una estabilidad considerable. Este dato tiene implicaciones importantes, como son:

1. que dichos reactivos sirvieron de "ancla" en las distintas versiones, evitando dispersiones inesperadas ya que, como mostraron los análisis de varianza, no ocurrieron diferencias significativas;

2. si en una aplicación subsecuente se deseara reducir el número de reactivos, sin sacrificar el valor informativo de la prueba, no habría duda de cuáles reactivos se podrían eliminar;

3. que si no se tratara de lograr fines experimentales, como en el presente caso, la calificación de los sujetos se debería hacer eliminando los reactivos desajustados, y

4. que esta calibración modificaría sustancialmente la calificación de los sujetos, atendiendo a la ponderación que podría darse a las respuestas según su grado de dificultad, en vez de tomar en cuenta, como dato ciego, al total de reactivos.

Una calificación basada en los tipos de reactivos que se resolvieron correctamente, además de ser más justa, también sería más realista, en la medida en que aumentaría considerablemente el poder de discriminación de la prueba, al hacer más evidentes las diferencias en el desempeño, vale decir de la habilidad de los sujetos.

Cabe reconocer que, en la larga historia de las técnicas para construir instrumentos de evaluación, esta propuesta no tiene nada de novedoso. Pero también cabe enfatizar que su operación correcta debe apoyarse, inevitablemente, en el uso de instrumentos cuyos reactivos estén calibrados con precisión, ya que los desajustes pueden generar resultados difíciles de interpretar, como pudo verse en el ejemplo de la figura VI.2, que ilustra algunos casos en que los sujetos podrían recibir una calificación que no reflejara su habilidad.

Si se siguiera el criterio de asignar una puntuación diferencial, por ejemplo 0 puntos a las respuestas incorrectas, 1 a los reactivos más fáciles, 2 a los reactivos medios y 3 a los de mayor dificultad, los datos de la figura VI.1, al convertirse a calificaciones ponderadas, tendrían los valores que se muestran en la figura VI.3.

Figura VI.3. Calificación ponderada

SUJ.	REACTIVOS									Aciertos	NIVELES DE HABILIDAD		
	1	2	3	4	5	6	7	8	9			CALIF.	
1	1	0	0	0	0	0	0	0	0	1	BAJO	1	
2	1	1	0	0	0	0	0	0	0	2		2	
3	1	1	1	0	0	0	0	0	0	3		3	
4	1	1	1	1	0	0	0	0	0	4		5	
5	1	1	1	1	1	0	0	0	0	5	MEDIO	7	
6	1	1	1	1	1	1	0	0	0	6		9	
7	1	1	1	1	1	1	1	0	0	7		12	
8	1	1	1	1	1	1	1	1	0	8	ALTO	15	
9	1	1	1	1	1	1	1	1	1	9		18	
Aciertos	9	8	7	6	5	4	3	2	1	TOTAL			
	FÁCIL			MEDIO			DIFÍCIL						
	NIVELES DE DIFICULTAD												

1 = Acierto
0 = Error

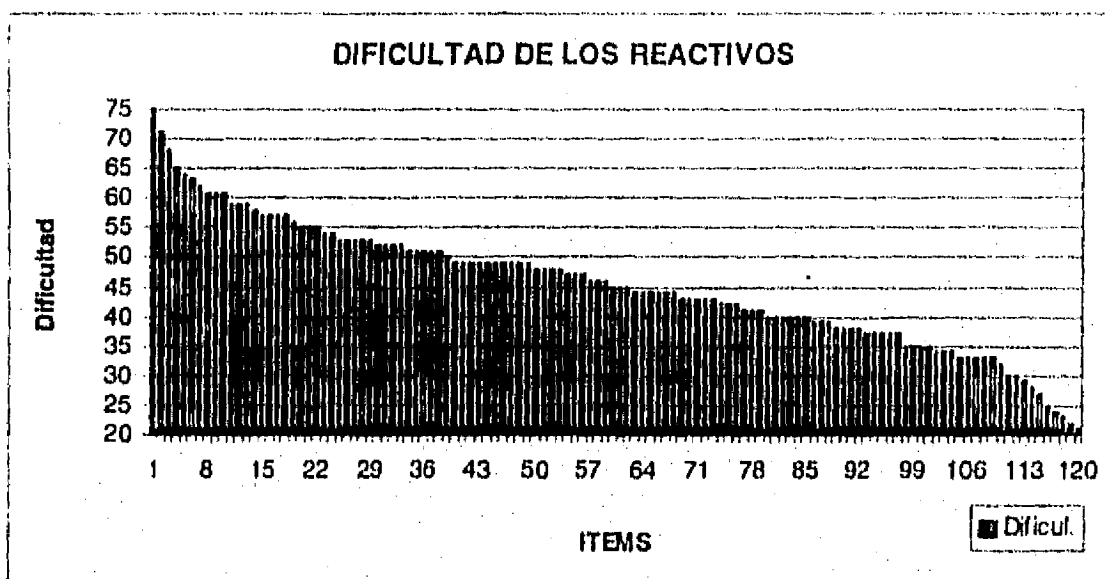
Los reactivos fáciles aportan 1 punto a la calificación, los medios 2 y los difíciles 3.

Otro punto a discutir es el desempeño general de los sujetos en la evaluación efectuada. La calificación escalada, que se generó a partir de los puntajes crudos, se extendió en un rango de 86 a 129. La habilidad baja tuvo un rango de 86 a 93 puntos; la media fue 94 a 106 y la alta de 107 a 129. Como ya fue señalado, los puntajes crudos fueron de 24 a 116, lo que indica una amplitud considerable; la habilidad de la muestra, estimada partir de la dificultad de la prueba, fue de -1.55 a 3.23. Los sujetos de habilidad baja estuvieron entre -1.55 y -0.72, la media estuvo en el rango de -0.68 a 0.64 y la alta de .68 a 3.23.

La gráfica VI.1 muestra la distribución de los 120 reactivos por grado de dificultad (menor a mayor), calculada a partir de las respuestas correctas (puntajes crudos). Puede advertirse que, excepto en los valores extremos, la distribución fue uniforme y sin cortes abruptos, en un continuo amplio.

El número de aciertos que los 300 sujetos alcanzaron en los 120 reactivos fue 18,075 de 36,000 posibles. Esto representa una proporción de 0.50 en relación a los sujetos y de 150.62 en relación a los reactivos. Si se advierte que en cada reactivo el máximo de aciertos posible es 300, se notará que el valor calculado de 150.62 constituye casi exactamente la mitad.

Gráfica VI.1. *Dificultad de los reactivos.*



La media de respuestas correctas obtenida por los sujetos, alcanzó el valor de 60.25 sobre 120 aciertos; el punto medio de la distribución marcado por la mediana, fue 54.50 y la calificación más frecuente fue 85. En relación a la estructura de la muestra, aproximadamente 200 sujetos formaron los subgrupos medio y alto, que en el continuo del número de aciertos va de 41 a 120.

En el caso de los reactivos, el rango de cada grado de dificultad fue dado por cortes de 40 aciertos, por lo que 80 ítems contuvieron los aciertos que definieron su dificultad como media y baja, conjuntando así el mayor número de respuestas correctas.

A continuación se presentarán varias tablas que resumen aspectos del desempeño de los sujetos y de los reactivos, como son la distribución de la dificultad de la prueba estimada con las puntuaciones crudas y con unidades convertidas a unidades escaladas por niveles de habilidad y de dificultad de las subpruebas, tipos de habilidad y de conocimiento. Estos

datos contribuyen a contextualizar la discusión de los resultados que se obtuvieron en la presente investigación.

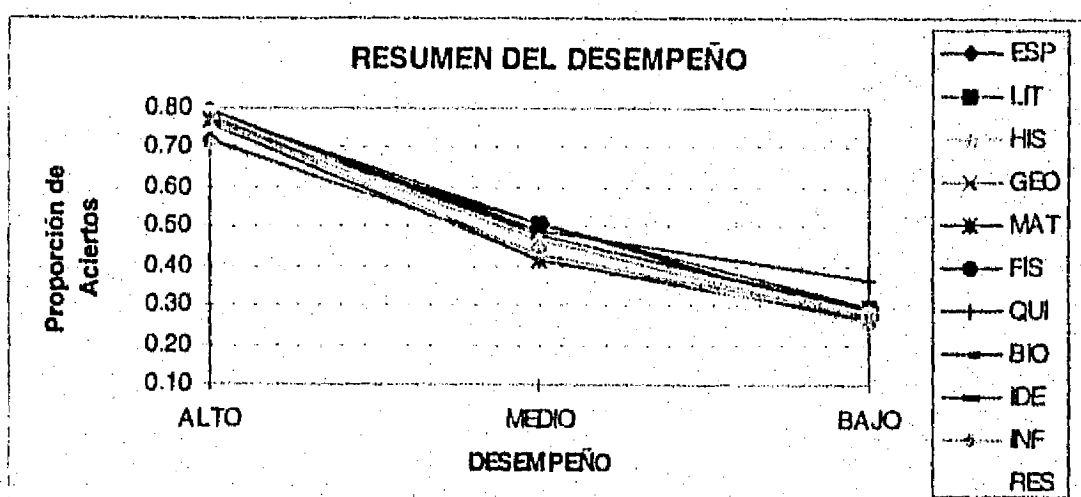
Tabla VI.3. Resultados en cada subprueba. Proporción de aciertos.

Habilidad	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO
ALTA	0.72	0.79	0.78	0.77	0.76	0.80	0.79	0.76
MEDIA	0.43	0.51	0.46	0.47	0.41	0.48	0.49	0.51
BAJA	0.26	0.30	0.26	0.29	0.27	0.29	0.36	0.27
Media	0.47	0.53	0.50	0.51	0.48	0.52	0.55	0.51

Tabla VI.4. Tipos de Habilidad y de Conocimiento. Proporción de aciertos.

Habilidad	IDE	INF	RES	Habilidad	CON	FAC	PRO
ALTA	0.78	0.71	0.79	ALTA	0.76	0.77	0.77
MEDIA	0.48	0.43	0.45	MEDIA	0.47	0.46	0.42
BAJA	0.29	0.26	0.28	BAJA	0.29	0.27	0.28
Media	0.52	0.47	0.51	Media	0.51	0.50	0.49
	HABILIDAD				CONOCIM.		

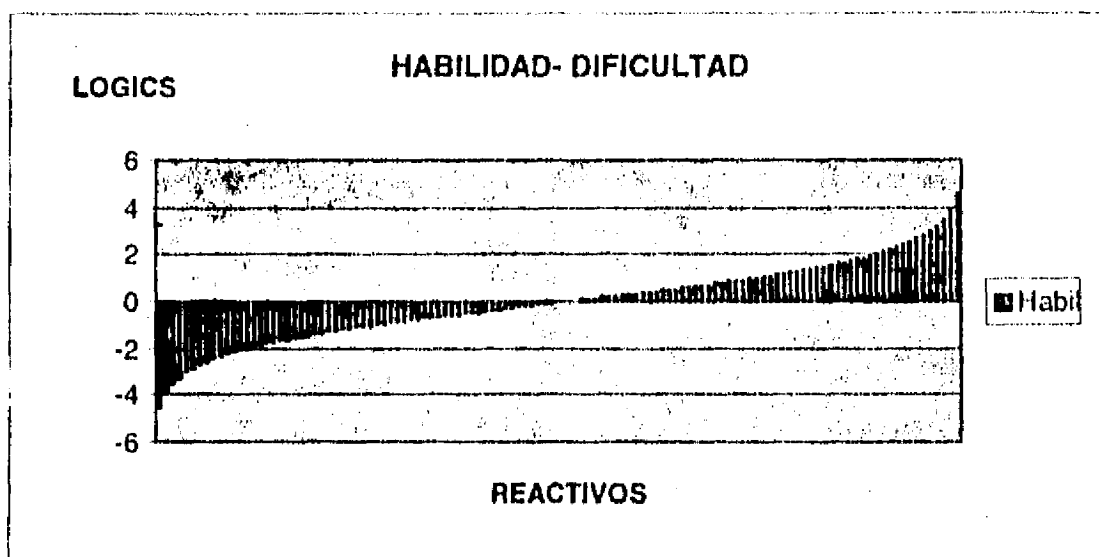
Gráfica VI.2. Desempeño de los sujetos en los tres indicadores.



Todos estos datos reafirman la congruencia de la relación entre el nivel de habilidad y el desempeño, derivada de los indicadores manejados en este estudio.

La siguiente gráfica ilustra otro aspecto de los resultados en los parámetros adoptados para identificar los perfiles de desempeño.

Gráfica VI.3. Distribución de la dificultad de los reactivos medida en unidades logic, asociada a la habilidad de los sujetos (n = 300).



Con base en estos resultados, puede concluirse que la prueba mostró estabilidad en la gran mayoría de los reactivos y que la distribución de las respuestas fue congruente con la estructura planeada, teniendo en cuenta que se aplicaron ocho subpruebas.

No obstante, fue notoria la tendencia a la adivinación y la existencia de trayectorias oscilantes, en algunos sujetos con desempeño bajo. El perfil general de estos sujetos indica patrones de respuesta difusos y dispersos.

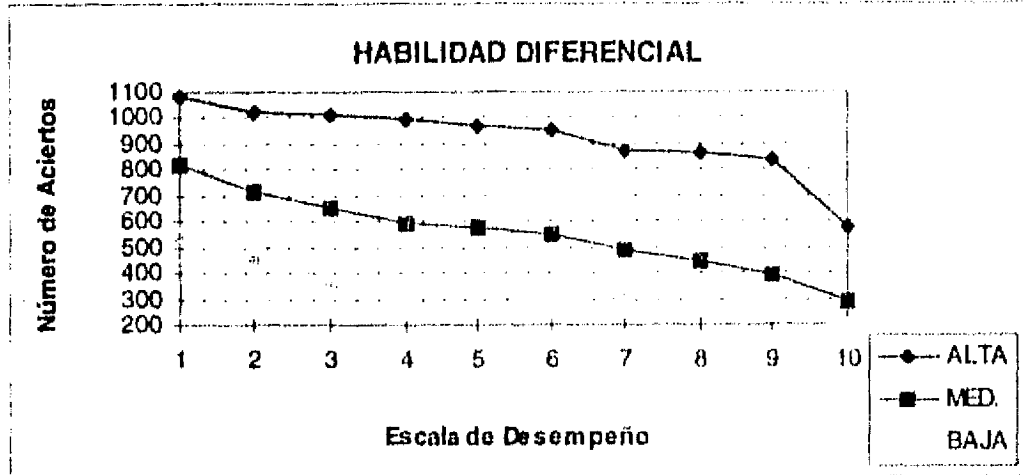
Para disponer de puntos de observación más amplios, acerca de las diferencias en los patrones de respuesta generados en el marco *habilidad-dificultad*, se construyó una escala de 10 grados de dificultad, cuyos límites estuvieron dados por cortes de 12 reactivos (10 % de la prueba).

De acuerdo con estos resultados, conforme aumentó la dificultad de los reactivos, disminuyó uniformemente el número de aciertos, tanto dentro de grupos como entre ellos.

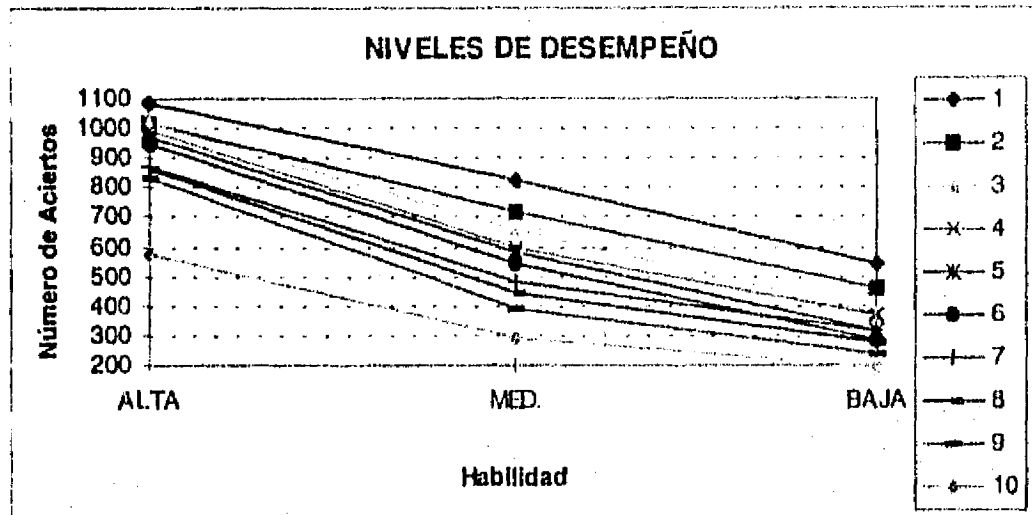
Por ejemplo, en el corte 1, que implicó el menor nivel de dificultad, hubo una cantidad de respuestas correctas mayor que en el 2, 3, 4, etc. que fueron progresivamente más difíciles. El grupo de mayor habilidad tuvo más aciertos que el grupo de habilidad media y éste que el de habilidad baja.

Las siguientes gráficas ilustran los resultados descritos en relación a la habilidad diferencial en la escala de 10 puntos, basada en la respuesta a reactivos de diferentes grados de dificultad.

Gráfica VI.4. Habilidad diferencial en una escala de diez puntos.



Gráfica VI.5. Niveles de desempeño por grupo de habilidad.



Al contrastar el peso relativo de cada una de las ocho asignaturas evaluadas en las subpruebas, de los tipos de habilidad y de los tipos de conocimiento es evidente la influencia diferencial de esas tres variables en la respuesta de los sujetos, además de las inherentes a quien responde.

Todo esto confirmó que, más allá de las particularidades de alguno de los grupos evaluados, fueron identificadas exitosamente las características propias de cada nivel de desempeño. El grupo *alto* se distinguió por ser apto en todas las subpruebas, siendo raro observar alguna inconsistencia en sus respuestas. El grupo *medio*, menos consistente, en las subpruebas mostró tendencias oscilantes, sin rebasar límites que van 10 puntos abajo de la media grupal y en algunas asignaturas tuvo un perfil más cercano al grupo alto. En los sujetos con rendimiento *bajo* el único elemento común fue la deficiencia generalizada.

A2. Discusión de los perfiles de desempeño asociados al contenido de la evaluación, tipos de habilidad y tipos de conocimiento.

Desde otro enfoque, mediante diversos análisis de regresión múltiple efectuados en SPSS (1993) y SYSTAT (1992) se procedió a la identificación de los ítems que tuvieran asociación alta con las asignaturas, las habilidades evaluadas y los tipos de conocimiento. La siguiente tabla resume los resultados principales.

Tabla VI.5. Resumen de resultados de las asociaciones entre los tres indicadores del desempeño.

Contenido	TOTAL DE REACTIVOS	HABILIDAD			CONOCIMIENTO		
		IDE	INF	RES	CON	FAC	PRO
ESPAÑOL	10	1	7	2	10	0	0
LITERATURA	8	5	1	2	5	3	0
HISTORIA	8	5	3	0	3	5	0
GEOGRAFÍA	8	7	1	0	5	3	0
MATEMÁTICAS	11	4	0	7	5	0	6
FÍSICA	8	1	1	6	2	2	4
QUÍMICA	8	5	2	1	6	1	1
BIOLOGÍA	8	7	1	0	6	2	0
TOTAL	69	35	16	18	42	16	11

Con base en la aproximación multicomponencial *IRT*, cabe afirmar que la probabilidad de que se acierte en un ítem, es función de factores tanto cognoscitivos como de la estructura del reactivo y de la tarea que demanda ejecutar, siendo todos ellos importantes y de naturaleza diferente. Esto quedó confirmado al analizar las asociaciones del contenido con la habilidad evaluada y el tipo de conocimiento implicado, aunque su peso varió en relación al nivel de desempeño.

A la luz de los resultados mostrados en la tabla, se aclararon algunos aspectos que conviene discutir. Por ejemplo, la competencia en *español*, o si se quiere, de las habilidades verbales, supone tener capacidades que no se limitan a la identificación y los hechos, ya que los reactivos con mayor peso en la evaluación de los sujetos se asociaron con inferir y manejar conocimiento de tipo conceptual.

En *literatura* predominó la habilidad para identificar, tal vez como un reflejo de la enseñanza poco significativa; *historia* se apoyó también en la identificación, sólo que en este caso se refiere a hechos; *geografía* requiere también identificar, pero es menos factual y más conceptual; matemáticas se

vinculó fuertemente con la habilidad de resolver y supone dos tipos de conocimiento: el conceptual y el procedural, lo cual es similar en física, sólo que en ésta participó también el conocimiento factual. *Química y biología* se asociaron a habilidades para identificar y al conocimiento conceptual.

Al analizar los resultados tomando al tipo de *habilidad* como base de la comparación, se observa que *identificar* estuvo presente con mayor peso en biología y con el conocimiento conceptual; *inferir* se asoció fuertemente con español y también con el conocimiento conceptual. Finalmente, la habilidad de *resolver* fue típica en matemáticas y se ligó a los conocimientos conceptual y procedural. Cabe subrayar la importancia para el desempeño académico que jugó el conocimiento *conceptual* sobre las otras variantes, confirmando la tendencia en la evaluación no tradicional a dejar de lado las tareas simples y el manejo memorístico del contenido del aprendizaje.

Tomando como eje al tipo de conocimiento, se advierte que el *conceptual* se liga particularmente a español y a la habilidad de inferir; el conocimiento *factual* estuvo presente en historia y su asociación más alta fue con la habilidad de identificar. Es claro que el conocimiento *procedural* es un componente básico del manejo matemático y que se acompaña de la habilidad para resolver.

Cabe señalar que en particular *español y matemáticas* tuvieron un peso importante en el desempeño de los sujetos, ya que aparecieron con los índices de dificultad mayor, como se mostró en la tabla VI.3. Esto fue similar en las habilidades y tipos de conocimiento, con la observación reiterada de que están asociadas al nivel de desempeño. Los sujetos del *nivel más alto* contestan correctamente los reactivos de mayor demanda cognoscitiva, como lo es la habilidad para hacer inferencias válidas a partir de la lectura y para resolver problemas aplicando procedimientos y principios adecuados, manejando competentemente hechos y conceptos relevantes.

El contraste con el *grupo más bajo* es importante, porque en este caso el desempeño es apenas suficiente en el manejo de hechos y tareas de reconocimiento, mostrando además inconsistencias importantes al resolver eventualmente alguna tarea de dificultad media y fallando en las de dificultad menor. La tabla VI.4 mostró un resumen del desempeño de cada nivel, con base en la proporción de aciertos.

Sin duda, el caso más complejo de analizar y el más preocupante, es el de los sujetos con rendimiento *bajo*, porque su perfil muestra en general un patrón menos definido. Para ilustrar los contrastes se utilizará la tabla VI.6, que contiene datos de 10 sujetos con diferentes desempeños.

De acuerdo con la proporción total de aciertos, los primeros cuatro sujetos son de nivel alto; los tres siguientes son de nivel medio y los últimos tres son de nivel bajo, indicando en cada caso tanto la suma como la proporción de aciertos.

Tabla VI.6. *Perfiles de desempeño diferencial.*

ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	SUM	PRO	ss
22	10	11	12	20	12	10	11	108	0.90	1
17	10	10	11	20	12	10	9	99	0.83	2
13	7	9	12	22	8	10	10	91	0.76	3
15	8	10	5	16	10	9	9	82	0.68	4
13	9	5	7	12	8	10	7	71	0.59	5
12	6	8	7	7	8	5	7	60	0.50	6
9	7	8	4	7	3	3	6	47	0.39	7
2	1	6	3	8	4	0	4	28	0.23	8
5	3	1	3	5	2	7	2	28	0.23	9
4	4	2	1	7	1	3	3	25	0.21	10

El sujeto 1 tiene puntuación alta en general y destacan matemáticas y español, únicas subpruebas con 42 ítems; en dos asignaturas logró 12/12, en dos 11/12 y en dos 10/12. El sujeto 2 está relativamente bajo en español (17/24), pero en todos los demás casos su desempeño es alto. Los sujetos 3 y 4 están también muy por encima de la media, resolviendo correctamente más de $\frac{3}{4}$ partes.

El sujeto 5 es claramente "medio alto" con un desempeño global de 71/120. En español y matemáticas está en 13/24 y 12/24, y en un caso (historia) está por debajo de la mitad de aciertos posible. En las demás asignaturas rebasa con holgura la relación de 6/12 aciertos. El sujeto 6 tiene en matemáticas 7/24 y en español 12/24, ubicándose apenas encima de la mitad de aciertos, con un total de 60/120. En tanto que el sujeto 7 es "medio bajo", con 47/120.

Al observar a los *sujetos de nivel bajo* puede advertirse que aún cuando dos de ellos tuvieron la misma calificación en la prueba (28/120), sus perfiles son diferentes, incluso uno obtuvo 0 aciertos en química. El sujeto 10, con el rendimiento más bajo (25/120) en español supera al sujeto 8 y casi lo iguala en matemáticas. Como ya ha sido señalado, el rasgo común que los caracteriza es la deficiencia generalizada.

Como conclusión global de la discusión de esta porción del análisis, cabe señalar que el enfoque *psicométrico* permitió alcanzar plenamente el primer objetivo general que fue propuesto para este estudio, en la medida en que:

1. Fue posible establecer con suficiencia los perfiles de desempeño en: a) el manejo del contenido temático, b) las habilidades escolares generales, que se asumieron como relevantes para el contexto instruccional y c) en los tres

tipos de conocimiento asociados al manejo conceptual de los materiales de aprendizaje.

2. Los resultados expresaron, desde varios ángulos, la consistencia de los indicadores que definieron el constructo, asegurando la validez requerida;
3. Los cinco grupos de comparación mostraron comportamientos altamente homogéneos, apoyando la confiabilidad de los resultados, y
4. El modelo teórico adoptado, que se basó en los principios psicométricos de la metodología IRT, cumplió su función orientadora y de sustento para alcanzar las metas de investigación propuestas.

A continuación se discutirán las implicaciones de los resultados que se obtuvieron al aplicar en este estudio la segunda estrategia metodológica prevista, en relación con el desarrollo y prueba de un sistema automatizado para el reconocimiento de los perfiles identificados en la fase psicométrica, que tomó forma en un neurocomputador.

B. DISCUSIÓN DE RESULTADOS: PORCIÓN NEUROCOMPUTACIONAL.

El análisis neurocomputacional permitió incursionar en una dimensión importante, que que rara vez es considerada y que es laborioso abarcar, como se muestra en el siguiente planteamiento.

Cuando se intenta precisar la ubicación de un sujeto evaluado, es necesario hacer una comparación con los indicadores de otros sujetos u otros subgrupos. Esto es, un perfil se estima como *alto* en comparación con otro que es de tipo *medio* y un tercero que es de tipo *bajo*. Si un sujeto tiene el 100 % de aciertos tendrá el perfil del más alto desempeño. Si alguien obtiene 0 % de aciertos también tendrá un perfil nítido.

Cuando se obtienen aciertos del orden de 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90/120, ¿qué puede concluirse? Sin duda se puede atender a la categoría dominante y clasificar al sujeto dependiendo del total de aciertos: entre 0 y 40 es *bajo*, entre 41 y 80 es *medio* y entre 81 y 120 es *alto*. Si un sujeto logra por ejemplo 85 aciertos, tiene un perfil alto, pero se parece mucho a los de desempeño medio y también en algunos aspectos a los de desempeño bajo. Un sujeto del grupo intermedio ¿es medio alto o medio bajo? La pregunta concreta es *cómo determinar en qué medida el perfil de un sujeto conjunta características de los tres niveles de comparación*, ya que conforme disminuye o aumenta el total de aciertos se empiezan a compartir rasgos de los otros perfiles.

La respuesta se encuentra en el *procesamiento neurocomputacional*, ya que señala claramente la composición del perfil de cada sujeto, indicando la categoría dominante y la proporción en que se es bajo, medio o alto, agregando así una dimensión de realismo que está ausente en la generalidad de las evaluaciones del desempeño.

En la tabla VI.7 se presentan de nuevo los datos de la tabla VI.6, agregando las columnas que indican los valores específicos del desempeño de los sujetos. Esta nueva información señala las particularidades de los perfiles (columna de *clasificación*) y la comunalidad entre ellos (columna de *categoría*).

Tabla VI.7. Perfil neurocomputacional del desempeño

Subpruebas								Totales		Clasificación			Categoría			ss
ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO	SUM	PRO	ALT	MED	BAJ	ALT	MED	BAJ	
22	10	11	12	20	12	10	11	108	0.90	0.991	0.000	0.009	1	0	0	1
17	10	10	11	20	12	10	9	99	0.83	0.988	0.001	0.011	1	0	0	2
13	7	9	12	22	8	10	10	91	0.76	0.979	0.004	0.020	1	0	0	3
15	8	10	5	16	10	9	9	82	0.68	0.914	0.072	0.086	1	0	0	4
13	9	5	7	12	8	10	7	71	0.59	0.589	0.919	0.420	1	1	0	5
12	6	8	7	7	8	5	7	60	0.50	0.453	0.992	0.547	0	1	1	6
9	7	8	4	7	3	3	6	47	0.39	0.456	0.969	0.544	0	1	1	7
2	1	6	3	8	4	0	4	28	0.23	0.000	0.000	0.999	0	0	1	8
5	3	1	3	5	2	7	2	28	0.23	0.000	0.000	0.999	0	0	1	9
4	4	2	1	7	1	3	3	25	0.21	0.000	0.000	0.999	0	0	1	10

Los primeros cuatro sujetos son claramente altos, indicado por el número 1 en la columna correspondiente a *categoría*. Sin embargo, el perfil del *sujeto 1* tiene un valor de 0.991 de perfil alto y 0.009 del perfil bajo, posiblemente porque los aciertos en literatura y química son inferiores a sus otras puntuaciones. El *sujeto 2* es 0.988 alto, 0.001 medio y 0.11 bajo. Se advierte que en este caso y en el de los *sujetos 3* y *4*, al disminuir el número y la proporción de aciertos se incrementan los valores de las categorías media y baja.

De los sujetos de categoría media, el *número 5* es "medio alto" (categoría 1,1,0) pues presenta perfiles alto (0.589), medio (0.919 como categoría dominante) y bajo (0.420), en contraste con el *número 6* que es "medio bajo" (categoría 0,1,1), aún cuando aparentemente está justo en el nivel medio (60/120 aciertos), tanto como el *sujeto 7* que tuvo 47/120 aciertos (categoría 0,1,1). Puede observarse también que los *sujetos 8, 9* y *10* son definitivamente bajos y que los valores asociados a su categoría de desempeño son 0,0,1.

A la luz de estos hallazgos cabe señalar también la similitud extrema que se observó al comparar los resultados de los modelos psicométrico y neurocomputacional en el reconocimiento de patrones. Los datos de cada caso mostraron diferencias mínimas y lo más destacable es que se logró reproducir los perfiles correspondientes a cada grupo cuando se tomó como

base *cualquiera* de los tres prototipos de neurocomputador que fueron contruidos para el presente estudio.

De igual forma, al comparar los parámetros de la población completa, se pudo observar la ausencia de diferencias en los perfiles, implicando que cualquiera de los tres prototipos contaban con una generalidad probada, para identificar con precisión los patrones de poblaciones mucho mayores.

Es cierto que las cinco muestras del estudio fueron muy similares, pero también lo es que la fidelidad del reconocimiento de los perfiles se mantuvo en un rango amplio de generalización. Como fue expresado en la porción de *método*, en las cinco versiones de prueba sólo 77 reactivos fueron comunes, lo que significa que el 36 % de los ítems fue diferente en cada caso. La relación es de 77 iguales contra 215 distintos, considerando 43 en cada una de las cinco versiones.

Puede concluirse que, dejando de lado los propósitos experimentales que orientaron este estudio, no es razonable asumir que para realizar una misma tarea se necesiten tres prototipos; ni que para que su funcionamiento sea apropiado se requieran muchas muestras de comparación homogéneas.

Por el contrario, una propiedad importante de la neurocomputación es su capacidad para trabajar, empleando un solo prototipo, con datos un tanto incompletos, inexactos o difusos, e incluso con aquellos para los cuales el investigador no tiene definido el patrón que emergerá del procesamiento. Partiendo de esta premisa, puede resaltarse su valor para el diagnóstico en el campo psicológico, en que con mucha frecuencia es necesario analizar información con características como las descritas.

También puede valorarse el por qué de su aplicación cada vez más amplia, para realizar tareas complejas en diversos campos científicos y tecnológicos, en que se requiere atender simultáneamente a gran número de variables y parámetros. La investigación psicológica contiene con una buena cantidad de problemas, cuyo tratamiento se beneficiaría, sin duda, con estas perspectivas innovadoras, productivas y confiables.

CAPITULO VII

SUMARIO Y CONCLUSIONES

El presente trabajo se enmarcó en líneas teóricas de la investigación psicológica cognoscitiva e instruccional y en lo metodológico en desarrollos de la psicometría y la simulación neurocomputacional, para explorar tres aspectos del desempeño en una evaluación diagnóstica, como base para establecer perfiles diferenciales: a) manejo de los contenidos temáticos de asignaturas de los programas de estudio del nivel de enseñanza medio superior, b) habilidades cognoscitivas y c) tipos de conocimiento.

La tesis sostenida en este trabajo, condujo a someter a prueba que la identificación de rasgos subyacentes a patrones de respuesta, inherente al diagnóstico psicológico, requiere en primer lugar del análisis psicométrico para establecer los parámetros de interés; y segundo, que el reconocimiento de dichos patrones puede hacerse confiablemente, usando procedimientos automatizados de alta precisión.

Para obtener evidencia experimental acerca de la fundamentación de dichos planteamientos, la identificación de patrones en la fase psicométrica se alimentó con los datos de la aplicación de un instrumento de medición del desempeño, similar a los que se emplean (CENEVAL, 1994) para evaluar conocimientos y habilidades, que se asumen como necesarios para ingresar al bachillerato. Los sujetos seleccionados para el estudio formaban parte de una población mayor a 20,000 egresados de secundaria, de los cuales se tomó al azar una muestra de 1500 sujetos, que fue desagregada en cinco grupos homogéneos, cada uno con 300. Para realizar la segunda fase del estudio, se utilizó un prototipo de neurocomputador, diseñado para la lectura y el análisis de las mismas respuestas procesadas en la fase psicométrica.

Los resultados confirmaron que las dos suposiciones que orientaron el estudio fueron acertadas, en la medida en que el procedimiento seguido en las dos vías permitió establecer exitosamente los perfiles de los sujetos y disponer de información útil para realizar diversos contrastes, con la meta de explorar los rasgos asociados a los niveles de habilidad diferencial.

En el capítulo inicial de este trabajo, fueron expuestos planteamientos teóricos y metodológicos relevantes a la investigación efectuada, que en conjunto contribuyeron a delinear el marco de referencia general y a ofrecer un panorama actualizado de los recursos y estrategias psicométricos que pueden ser aplicados al análisis e identificación de patrones de respuesta.

Con esa base, en la planeación y desarrollo del estudio se prestó atención a diversos problemas metodológicos, tanto conceptuales como de instrumentación, que suelen manifestarse en estudios como el efectuado ya que, cuando son ignorados, tienden a afectar sensiblemente la confiabilidad y validez de la evaluación. Entre estas debilidades, fueron resaltadas las que tienden a asociarse con factores que vulneran la validez de constructo, es decir, la representación apropiada de la o las variables que interesa medir en una prueba.

Este problema fue subsanado con suficiencia, al contar con una teoría psicológica orientada cognoscitivamente a la investigación de problemas instruccionales, de la que fueron derivados los componentes de un modelo de evaluación que, además de ofrecer soporte a la planeación y desarrollo del estudio, proporcionó guía para la selección de las variables relevantes. Igual peso tuvo en la calidad del estudio el observar la multidimensionalidad de los fenómenos que se abordaron al evaluar el desempeño escolar.

Estos antecedentes llevaron a reconocer la necesidad de adoptar un enfoque integrador en este trabajo, de manera que la tarea de evaluación no se viera contaminada al privilegiar algunos componentes del desempeño, por ejemplo el manejo del contenido temático, ignorando o minimizando otros (por ejemplo, las habilidades escolares). Las variables seleccionadas han recibido sustento teórico y empírico de estudios realizados por Johnston (1989), Pozo (1990), De Vega (1989), Bransford (1979), Gagnè (1995) y Anderson (1983), además de trabajos propios realizados en el posgrado de la Facultad de Psicología de la UNAM: Castañeda y López (1988a y 1988 b; 1990 y 1991), Castañeda, López y Romero (1987), Castañeda, López, Gómez, Cabrera y Orozco (1989), López y Castañeda (1990, 1992), López, Castañeda y Gómez (1989) y López, Castañeda, Pineda y Orduña (1992).

Desde el punto de vista psicométrico, fue relevante tomar como base recomendaciones extraídas de los trabajos de Embretson (1983, 1984, 1985, 1991) y de Wrigth y colaboradores (Wrigth, 1977; Wrigth y Masters, 1982 y Wrigth y Stone 1979). Entre los puntos de coincidencia con estos autores, caben destacar los siguientes: 1. el conceptualizar como *atributos o rasgos latentes* las variables psicológicas a medir y 2. incorporar el modelo *IRT* (Item Theory Response), como marco metodológico general.

En el diseño de la prueba se incluyeron tres habilidades cognoscitivas generales (identificar, inferir, resolver), necesarias para el trabajo escolar y tres tipos de conocimiento (factual, conceptual, procedural), comunes en los campos temáticos del contexto escolar. Todo esto permitió disponer de una taxonomía de variables bien definidas, para ser usadas como indicadores del desempeño: contenido, habilidades y tipos de conocimiento.

Los contenidos de las subpruebas fueron seleccionados del temario de estudios vigente (SEP, 1994), con el apoyo de profesores en servicio y se

siguió un formato de examen estándar en nuestro medio, que se utiliza para evaluar el ingreso a los estudios de bachillerato (CENEVAL, 1994).

Las variables exploradas se seleccionaron a partir del modelo teórico que fue descrito en el capítulo I. El estudio realizado mostró que todas ellas estuvieron relacionadas consistente y significativamente con el desempeño académico diferencial y que contribuyeron a la configuración de los patrones de interés, con rasgos observables y medibles. Por lo tanto, fue posible detectar perfiles que permiten inferir factores de riesgo potencial para el éxito en los estudios subsecuentes, siendo de relevancia particular los que invariablemente se vincularon con el desempeño bajo.

Como fue planteado al definir el problema de investigación abordado en este estudio, existe una carencia grave de información relevante, acerca de conocimientos, habilidades, actitudes y aptitudes que el sistema escolar está desarrollando en los alumnos de todos los ámbitos educativos en nuestro país. Esto es importante, porque implica que se ignoran aspectos claramente vitales, que deberían ser fuentes indispensables para promover cambios curriculares, seleccionar contenidos de estudio, prescribir prácticas remediales. En suma, para el desarrollo de *habilidades de pensamiento* (de Sánchez, 1992; Herrnstein, Nickerson, de Sánchez y Swets, 1986).

Los factores que se asocian al desempeño escolar en los cursos, sin duda tienen nexos fuertes con variables que van más allá de la organización familiar y los perfiles socioeconómicos, porque no pueden ser ajenos a los contextos instruccionales en que se asume que el aprendizaje tiene lugar. Por sus implicaciones, es prioritario detectar los que afectan negativamente el aprendizaje y favorecen la deserción escolar.

El segundo objetivo general propuesto para este estudio, se centró en la necesidad de apoyar a la investigación y práctica psicológicas, usando sistemas automatizados con capacidad probada para realizar, en una gama amplia de modalidades y aplicaciones, las tareas especializadas que buscan identificar y reconocer confiablemente patrones de respuesta.

La presentación del enfoque neurocomputacional al reconocimiento de patrones de respuesta, abordada en el segundo capítulo, contuvo una exposición amplia, que buscó fundamentar y contribuir a la formación de un marco de referencia, para contextualizar la utilización de esta tecnología en la investigación psicológica.

En el capítulo mencionado, se expresó que el procesamiento basado en la neurocomputación es una derivación tecnológica innovadora, que comparte características del dominio psicológico (corriente cognoscitiva) y del campo de la Inteligencia Artificial, que tomó su nombre de la analogía con la forma en que los sistemas biológicos transmiten información, por

ejemplo en las estructuras del sistema nervioso central, al valerse en ambos casos de redes para la propagación de los impulsos o señales.

Una diferencia importante entre las redes neurodales (simuladas) y las neuronales (biológicas), es que las primeras se diseñan para que cumplan funciones que van más allá de la transmisión de información, ya que pueden ser *entrenadas* para que *aprendan* los tipos de relación fijados como meta y *recuerden* con exactitud los valores o parámetros aprendidos, de manera que al recuperar la información lo hagan con niveles de error muy bajos.

El uso de estos términos, que son propios del campo psicológico, en el contexto de la neurocomputación están bien definidos. El *aprendizaje* es concebido asociativamente como una adaptación o modificación de pesos en las conexiones, en respuesta a estímulos presentados como entradas (por esta razón, estos entornos han sido llamados sistemas conexionistas). El proceso por el cual una red aprende es visto como un *entrenamiento*. Los estímulos que se presentan constituyen *respuestas deseadas* o esperadas a dichas entradas de información y el *recuerdo* es una tarea de recuperación de información.

En este contexto, la segunda fase del estudio se enfocó a identificar, en los mismos grupos de comparación que fueron utilizados para el análisis psicométrico, los patrones de respuesta que definieron los tres perfiles de desempeño escolar (alto, medio y bajo), para proceder a su reconocimiento en forma automatizada, aplicando la simulación neurocomputacional. Esta tecnología demostró ser idónea para realizar la detección de rasgos diferenciales, condición que es indispensable en las tareas de diagnóstico psicológico.

Para hacer más objetiva la presentación de los resultados, se ilustraron todas las fases del procesamiento, iniciando con información de las matrices de entrada de datos y del entrenamiento, se incluyó una porción del listado de corrida de datos y las pantallas de identificación de patrones de desempeño alto, medio y bajo. Se presentaron también ejemplos de matrices de salida y prueba, para concluir con la matriz final que contuvo la identificación de los patrones de desempeño.

El prototipo de neurocomputador construido para realizar este estudio, fue entrenado para aprender a reconocer los tres patrones de desempeño identificados en la fase psicométrica, de manera que al presentarle los datos de sujetos (calificaciones) distintos a los empleados para el aprendizaje, indicó automáticamente la categoría correspondiente. Cuando al prototipo se le presentó un conjunto de datos nuevos, la red recordó lo aprendido y lo aplicó a evaluar la nueva información. La primera etapa (entrenamiento) se logró con un margen de error menor a 0.10 %; la tarea de reconocimiento se produjo con cero errores.

El primer neurocomputador fue entrenado con las respuestas de los sujetos del grupo 1, luego se aplicó a reconocer los patrones de los grupos 3 y 5; lo mismo se hizo con el prototipo 3, tomando los grupos 1 y 5 para prueba; y en el grupo 5 acerca de los grupos 1 y 3. El tiempo que se requirió para entrenar un prototipo de neurocomputador fue apenas mayor a cinco minutos; una corrida de prueba con 240 sujetos tardó un minuto; reconocer el perfil de 1500 sujetos se hizo en menos de dos minutos.

Los resultados mostraron una similitud extrema al comparar los datos de los modelos psicométrico y neurocomputacional al reconocer los patrones. Los datos de cada caso tuvieron diferencias mínimas y lo más destacable es que se reprodujeron los perfiles correspondientes a cada grupo cuando se tomó como base *cualquiera* de los tres prototipos de neurocomputador construidos para el presente estudio: uno para el grupo 1, otro para el 3 y un tercero para el 5, con el objeto de poder contrastar su funcionamiento.

En una corrida alterna y por separado de cada prototipo, aplicada a los sujetos del total de la población abarcada, con $n=1500$ al sumar los cinco grupos, se conservó la precisión y homogeneidad de cada neurocomputador al generar perfiles que permiten identificar a los 500 sujetos de desempeño alto, a los 500 de desempeño medio y a los 500 de desempeño bajo y conocer en cada caso el peso proporcional de cada categoría. Un dato clave es la respuesta en las ocho subpruebas, ya que contribuyó a delinear el perfil integral del desempeño de cada sujeto.

Una confirmación adicional de las capacidades del prototipo se derivó de probar una forma alternativa para categorizar sujetos que habían resuelto ítems construidos en el contexto de tareas piagetianas para identificar estadios de desarrollo cognoscitivo. Se utilizó un prototipo neurocomputacional para realizar las tareas de reconocimiento de patrones de una submuestra de 62 sujetos que respondieron 43 reactivos. La lectura de estos datos y la elaboración de la matriz final le requirió al prototipo un minuto con 8 segundos.

En conclusión, los resultados del estudio ofrecieron evidencias claras a favor del procesamiento automatizado para analizar patrones de respuesta e identificar perfiles de interés. Las implicaciones son muy importantes, desde varios puntos de vista. Una de ellas es la capacidad probada del prototipo construido, cuyo diseño relativamente simple mostró que un entrenamiento mínimo (240 casos) fue suficiente para categorizar, sin errores, a una población compuesta por 1500 sujetos.

Aún cuando no fue un objetivo del estudio atender a otros indicadores fuera de los descritos, es plausible establecer reducciones en la extensión de los instrumentos a partir del análisis de Rasch para calibrar los reactivos

y de los de regresión lineal para identificar los ítems que aportan el mayor peso de la varianza, digamos el 90 %. De esta manera se puede conservar la potencia de la prueba con un mínimo de reactivos. Si el entrenamiento neurocomputacional fuera hecho con los parámetros de esta selección, tal vez sería necesario entrenar a un prototipo con sólo 60 reactivos, para tener éxito en las tareas de reconocimiento.

El nombre de este método encubre el valor generalizado de su utilidad real, pues fue desarrollado, en el marco de la Inteligencia Artificial, con el único propósito de disponer de un sistema que fuera capaz de *aprender* a identificar y reconocer patrones, rasgos, características y tendencias, que en sentido estricto no tienen nada que ver con el campo neurológico, ni con el psicológico, en tanto que su uso no está circunscrito a alguna disciplina en particular. Su aplicación se ha vuelto común en materia de informática, medicina, ingeniería, finanzas, etc.

Una duda razonable, que ha estado presente en este trabajo, se refiere a la calidad de los *constructos* con que, en general, se busca medir el desempeño, porque las teorías psicológicas y los modelos metodológicos tienen todavía mucho camino por recorrer en la exploración de procesos tan complejos y elusivos como los cognoscitivos. La capacidad para medir el cambio sigue siendo un reto difícil de resolver, por lo que a la fecha la evaluación sigue siendo estática e incompleta.

No obstante estar ya de cara a un nuevo siglo, hay aspectos que pese a su extrema longevidad, siguen ofreciendo escollos a la investigación que se realiza en varias áreas de nuestra disciplina, mostrando que lo aprendido es insuficiente para alcanzar claridad en la comprensión de fenómenos que se antojan comunes, por su frecuencia y por su ocurrencia. Las discusiones interminables sobre competencias, aptitudes, habilidades y destrezas son un buen ejemplo de confusión conceptual, tanto como las que aluden a la ejecución, el rendimiento y el desempeño.

Es incierto el poder determinar cuánto de esta lentitud del crecimiento de nuestra disciplina proviene de las condiciones que han prevalecido en la educación, desde la formación básica hasta la universitaria, en comparación con las que existen en naciones y centros educativos de primer nivel en el plano internacional. Pero hay también una porción que debe asociarse a la resistencia al cambio para adoptar nuevas líneas de pensamiento, nuevos métodos de análisis y de representar los objetos de estudio.

Tal vez por esa razón las palabras escritas por Díaz-Guerrero en 1966 conservan su vigencia:

“La posibilidad de aprendizaje en hombres y máquinas viene a ser uno de los elementos más importantes en la batalla que libra el hombre ante la

tendencia natural del universo al desorden, a la desorganización y a la confusión, es decir, a la entropía" (página 32).

Cuando Negrete (1990) adoptó para una obra pequeña en extensión, pero muy rica en implicaciones el título: *Inteligencia, aunque sea Artificial*, puso la mira en una realidad nacional, no en una fantasía de la ciencia ficción. Y debe decirse que esa fue la consigna adoptada en 1956, hace ya 40 años, en la *Primera Reunión Internacional de Inteligencia Artificial*. A ella acudieron físicos, matemáticos, filósofos, lingüistas, ingenieros y psicólogos, destacando por su número el grupo de investigadores japoneses. En este evento se discutió acerca del uso potencial de las computadoras, en la simulación de cada uno de los aspectos del aprendizaje, o cualquier rasgo de inteligencia.

¿Cómo participar en estos desarrollos, si no se tiene inclinación para compartir una visión interdisciplinaria de los fenómenos que se estudian?

En México, una línea de crecimiento intelectual que entonces parecía promisoría, la llamada *tecnología educativa*, se empantanó en supuestos que parecían consistentes en los años sesenta; después vino otra forma de tecnología, la *ola computacional*, ofreciendo un sin número de atractivos para desarrollar el aprendizaje en las escuelas y ahora, con la moda de la *multimedia*, parece estar tomando un nuevo aire. Estas posibilidades han convertido a nuestro país en un extenso mercado de consumo y, en la mayoría de los casos, no se ha superado el ser, tan solo, buenos usuarios de *software*. Pero ¿cómo acelerar el avance en la productividad intelectual? y ¿cómo mejorar la calidad del proceso educativo en México, sin un desarrollo tecnológico propio, orientado a fortalecer la formación de los estudiantes en todos los niveles educativos?

Una opción expuesta y reiterada en el presente trabajo, señala que no es aceptable que se sigan ignorando aspectos elementales de la formación de los alumnos, ya que no se nota, siquiera, la intención de explorarlos. Si se entendiera el significado de esta carencia, podría acudir a desarrollos tecnológicos, como el aplicado en esta investigación, para atenderla.

Para cerrar este capítulo, conviene hacer un recuento de los hallazgos principales en los resultados del análisis psicométrico y neurocomputacional.

1. Los cinco grupos de sujetos usados para la comparación ($N=1500$; $n = 300$), fueron inicialmente homogéneos en su estructura y composición, ya que al compararlos entre sí, tomando como base los aciertos totales de cada uno en las tres variables de interés: a) subpruebas, b) habilidades y c) tipos de conocimiento, en ningún caso las diferencias fueron significativas.

Este resultado ofreció un soporte imprescindible al estudio, permitió tomar al azar uno de los cinco grupos iniciales y hacer análisis de mayor

extensión y profundidad, evitando así redundancia innecesaria. La porción neurocomputacional, basada en los resultados de los grupos 1, 3 y 5 reprodujo dicha homogeneidad.

2. Cuando se procedió al análisis de los tres niveles de desempeño, a manera de subgrupos de habilidad: alta, media y baja, se observaron con claridad las diferencias *entre ellos* pero no dentro de ellos, indicadas por los análisis de varianza del total de aciertos, coincidiendo los resultados en las tres variables: subpruebas, habilidades y tipos de conocimiento. Esto fue así porque hay una influencia muy grande de esas variables en la respuesta del sujeto, además de las que son inherentes a quien responde.

3. Los resultados afirmaron que existen características específicas en cada nivel de desempeño y que éstas pueden ser identificadas con cierta precisión. El grupo de mayor habilidad se distinguió por ser exitoso en todos los aspectos, mostrando consistencia en sus respuestas, salvo cuando los reactivos estuvieron mal contruidos, como pudo advertirse en dos casos. El grupo de desempeño *medio* tuvo altas y bajas, sin rebasar límites cercanos a la media grupal; en algunas asignaturas su perfil se acercó al del grupo alto. El grupo *bajo* mostró un perfil errático, inconsistente y con tendencia a la adivinación. Sus debilidades fueron obvias en los tres indicadores aplicados.

4. El estudio permitió identificar también, mediante diversos análisis de regresión múltiple, los reactivos que tuvieron mayor asociación con las asignaturas, habilidades y tipos de conocimiento que afectaron de manera significativa el desempeño en la evaluación. Por esta razón, contribuyeron a la definición de los perfiles diferenciales, ofreciendo información valiosa para fines curriculares.

5. El nivel de análisis que se deseó destacar en esta investigación fue primordialmente el de la habilidad de los *sujetos* y para favorecer la visión integral de su desempeño pudieron delinearse los perfiles con base en:

a) el total de respuestas correctas y la proporción de aciertos en toda la prueba, lo cual permitió ubicar a cada subgrupo en un rango de ejecución (alta, media o baja); b) el total de respuestas correctas y la proporción de aciertos en cada subprueba (biología, español, física, etc.) e incluso en los contenidos del temario relacionados con cada asignatura, atendiendo a la clasificación de cada ítem; c) el total de respuestas correctas y la proporción de aciertos en cada tipo de habilidad; d) el total de respuestas correctas y la proporción de aciertos en cada tipo de conocimiento; e) el nivel de habilidad, escalado a partir de la dificultad de los reactivos; f) la calificación escalada, calculada en relación al patrón de respuestas en toda la prueba, y g) el percentil en que se ubicó cada sujeto, en relación a la muestra de referencia.

6. Las especificaciones del estudio, en relación al tipo de evaluación y las características de los sujetos, fueron motivadas por el interés de mostrar, en un aplicación concreta y relativamente simple, la capacidad disponible para apoyar al diagnóstico psicológico automatizado, mediante desarrollos tecnológicos apropiados. Los resultados indicaron que la simulación neurocomputacional cumplió esos requisitos, al proporcionar herramientas de trabajo flexibles para adaptarse a la naturaleza de problemas comunes en nuestra disciplina y aportar soluciones viables e inteligentes.

7. La evidencia adicional obtenida en este estudio, al proceder con éxito a la clasificación de sujetos que respondieron reactivos basados en cuatro categorías piagetanas, es un ejemplo claro de la flexibilidad de estos entornos neurocomputacionales. Otros desarrollos en curso están dirigidos a identificar los perfiles diferenciales de los estudiantes de las cuatro áreas disciplinarias, que agrupan el nivel de enseñanza superior. También se está trabajando en un prototipo para realizar tareas complejas, basadas en la predicción, para abordar otros problemas que son socialmente relevantes, relacionados con estilos de vida y factores predisponentes a trastornos tan graves como el cáncer y el SIDA.

8. En conjunto, los resultados confirman sólidamente la capacidad de los modelos teórico, psicométrico y neurocomputacional adoptados, para dar cuenta y con suficiencia del desempeño académico diferencial, al permitir incorporar desarrollos y estrategias generadas en la investigación, que van más allá de las particularidades de la muestra y grupos estudiados. Los datos aportaron una cantidad considerable de evidencia empírica en su favor y señalaron algunos problemas en la instrumentación que, si bien ameritan revisión, no interfirieron con los resultados presentados. Más bien, fueron útiles para ilustrar de qué manera pueden conducir a conclusiones poco fundamentadas.

9. En general, el valor psicoeducativo de este estudio se desprende del aporte de información amplia y sustancial, útil para delinear marcos descriptivos y explicativos de fortalezas y debilidades en la formación de los estudiantes, en el presente caso de los que cursan niveles educativos medio superior y superior. El interés por contar con información relevante y completa, partió de la posición que se adoptó al plantear el problema de investigación: *un perfil completo y realista puede establecerse sólo cuando se tiene una visión integrada y multidimensional de las variables que configuran su estructura.*

10. La confiabilidad del reconocimiento de patrones quedó establecida por la convergencia al 100 % de las lecturas de los tres neurocomputadores, actuando como jueces independientes. Al no haber diferencias se demostró experimentalmente la calidad del diseño y de su implementación.

Las implicaciones psicoeducativas de los resultados antes descritos, viendo más allá de los límites de la presente investigación, pueden ser analizadas a la luz de las siguientes consideraciones.

Prácticamente desde sus inicios, la psicometría ha adoptado dos supuestos que recibieron amplia aceptación: 1°. los humanos diferimos unos de otros y 2°. las diferencias son significativas y medibles. En lo que se refiere a la existencia de diferencias entre los individuos, tal concepción se han manifestado en una gama muy amplia de distinciones. Por ejemplo, numerosos aspectos relacionados con patrones conductuales, habilidades cognoscitivas, aptitudes para realizar tareas físicas y mentales, temperamento e intereses, contrastes entre géneros y atributos en los rasgos culturales, como creencias, valores y actitudes.

En lo que toca a la forma en que dichas diferencias pueden ser medidas, Sternberg (1992) se vale de una analogía para expresar que es muy poco lo que ha cambiado después de un siglo de evaluación psicológica. Señala que, desde que fue creado el primer automóvil, pocas cosas han cambiado. El primer vehículo, que hoy se antoja primitivo, lento e incómodo, inició una empresa de gran éxito comercial, porque encajó perfectamente en una industria manejada por el mercado. Esta historia parece ser similar a la ocurrida en la evaluación de la inteligencia. Desde los primeros trabajos de Binet y Simon hasta nuestros días, en el campo de la evaluación de habilidades han habido cambios notorios en la construcción de pruebas, pero todas siguen los primeros diseños. Las pruebas de hoy están mejor presentadas, sus ilustraciones las hacen más atractivas, se administran más fácilmente, dan más información en menos tiempo y pueden ser efectuadas y calificadas por una computadora.

Si se deseara hacer un recuento de los mayores logros en la medición educativa, se haría necesario reportar los aportes tanto de la aproximación "tradicional", como la de los avances "no tradicionales", aspecto que se abordará más adelante. Puede imaginarse que al intentar dar cuenta de los fundamentos psicológicos y psicométricos que han adoptado las dos vertientes, se encontraría un rasgo común: teorías largas, densas, complejas y, también, insuficientes e inacabadas, junto con algunas características generales que pueden identificarse por sus propósitos prácticos en el desarrollo de instrumentos, entre las cuales cabe citar:

1. Que las pruebas tengan valor predictivo. El uso principal de las pruebas continúa siendo la predicción de algún rasgo, por ejemplo: rendimiento escolar.

- 2o. Que las pruebas sean confiables. Se espera que los puntajes sean relativamente estables para que sean útiles.

- 3o. Que las pruebas sean válidas. Este requisito requiere que se demuestre que la prueba mide aquello para lo cual fue hecha.

4o. Que se guarde alguna correlación con los resultados que ofrecen otras pruebas similares.

5o. Que las pruebas puedan administrarse con facilidad, por ejemplo, por profesores.

6o. Que las pruebas sean de interpretación fácil, de manera que los resultados puedan ser comunicados con facilidad y comprendidos por personas no especializadas.

7o. Que la calificación de las pruebas sea objetiva, para evitar sesgos en la interpretación y ambigüedad en los resultados.

8o. Que ofrezcan un nivel atractivo de la relación costo-beneficio.

Aunque esta lista de criterios no es exhaustiva, ni aplicable a todas las instancias en que se manejan pruebas psicológicas, es representativa de los aspectos más comunes en la construcción de instrumentos.

Por supuesto, debería también hablarse de los marcos de referencia teóricos y metodológicos que se han adoptado para fundamentar la medición. En este sentido, ha habido varios señalamientos. Algunos de ellos indican que existe un rompimiento entre la tecnología de elaboración de pruebas y la teoría psicológica. Otros sostienen posiciones anti-pruebas, aludiendo a que el problema central está en el manejo que se hace de las pruebas, asumiendo que se toman como base para tomar decisiones sin contar con un respaldo válido. De igual manera, se ha considerado, también, que la construcción y uso de las pruebas responde a las presiones del mercado, dejando de lado, por ejemplo, un interés real en la formación de los estudiantes.

La teoría llamada "tradicional" en la medición psicoeducativa, ha sido señalada por sobresimplificar y distorsionar la concepción del aprendizaje humano y por asumir que el potencial para el aprendizaje puede ser evaluado usando instrumentos unidimensionales, estáticos, y culturalmente neutros, a la luz de una medición marcadamente de tipo correlacional. En la aproximación "no tradicional", ahora se está haciendo muy común hablar de evaluación adaptativa, calibración de reactivos con métodos psicométricos y parámetros múltiples, desarrollo de tecnología asistida computacionalmente para construir instrumentos multidimensionales que se apoyan en teorías psicológicas que defienden la multicomponencialidad de los modelos, etc.

Sin embargo, junto con estos logros, que se presumen trascendentes, es también común en nuestros días observar algunas debilidades muy gruesas, señaladas reiteradamente acerca del diseño y uso de instrumentos de evaluación, sin que parezca advertirse la gravedad de cometerlas y, a la vez, las implicaciones que su uso inadecuado traen consigo. Entre ellas se encuentran:

1o. El diseño o plan de la prueba, que define qué, cómo, a quién y con qué medir, es uno de los pasos críticos que, como regla, se omite por las instituciones. Por el contrario, el examen de "aportación de reactivos" se ve muy favorecido, al creerse que si los reactivos tienen diferentes fuentes de procedencia y luego se juntan, la prueba será más completa y más representativa.

2o. Omisión de estudios piloto y de validación externa, previos a las aplicaciones formales, que permitan ajustar los instrumentos y los reactivos, hacer adaptaciones en las instrucciones, determinar el tiempo real que se requiere para responder y, sobre todo, para determinar las características psicométricas de los ítems y calibrarlos. Hay una cierta tendencia a creer que los reactivos se deben calibrar hasta después de que han sido aplicados, dejando fuera toda opción de beneficio para la evaluación actual. También lo es el creer que los reactivos se pueden calibrar con independencia a las respuestas de los sujetos.

3o. Después de que se aplican las pruebas se califican las respuestas y se difunden los resultados. Luego se hace un análisis detallado de los reactivos, para determinar sus niveles de dificultad y de discriminación, mediante alguna técnica que, por lo general, no explora algún indicador adicional. Lo grave es que, antes de dar calificaciones, no se remueven los reactivos defectuosos, que por su peso podrían incluso cambiar dichos resultados.

4o. No obstante que se asume que los reactivos difieren en dificultad y en complejidad, ya sea por su contenido, por la habilidad que miden, etc., se asigna el mismo valor a todos los ítems, haciendo imposible distinguir los niveles diferenciales en el desempeño, excepto por el indicador grueso que constituye la simple suma de aciertos.

5o. Es muy común que la dificultad de una prueba se represente como el porcentaje de sujetos que responde correctamente un ítem. Sin embargo:

a) El hecho de no resolver correctamente un reactivo no dice nada sobre la causa del error, de manera que no es posible saber si la frecuencia de respuestas erróneas de un ítem estriba en la dificultad del concepto o en otro factor, como instrucciones confusas, lenguaje rebuscado, contenidos que no corresponden a los temarios o a los planes de estudio, falta de tiempo para responder, etc.

b) El porcentaje de respuestas correctas no toma en cuenta la elección al azar, por lo que una parte importante de la varianza de error puede tener como fuente la respuesta aleatoria, reduciendo así la confiabilidad de la prueba.

c) La escala que se basa en el porcentaje de aciertos no puede considerarse como una escala de unidades equivalentes de dificultad. Dicho

de otra manera, carece de una unidad de medida, que exprese los niveles correspondientes a fracciones iguales de la escala, para la cual se diseñaron los reactivos.

El problema de la carencia de escala de unidades equivalentes, ya sea de dificultad o de otro tipo es el más complejo y recurrente en la medición psicológica en general y en la educativa en particular. Si se valora lo que aporta el análisis de Rasch, puede advertirse que el proceso de calibración de reactivos representa una opción interesante, en la medida en que su aplicación considera tanto los aciertos como los errores de cada sujeto o grupo de comparación, traduciéndolos a calificaciones escaladas y expresándolas como probabilidades asociadas las trayectorias de respuesta en toda la prueba.

Al relacionar el nivel de dificultad con el de habilidad de la persona, se establece que cuando el nivel de dificultad de un ítem y el nivel de habilidad son iguales, entonces la probabilidad de que un miembro de ese grupo falle u obtenga un acierto son iguales, al nivel de 50 %.

La escala estándar resultante de la diferencia entre la dificultad y la habilidad se expresa en límites que van de +3.00 a -3.00 y es factible establecer su equivalencia con la proporción de aciertos en la prueba. Por ejemplo +3.00 se traduce a una proporción de 95; +2.00 en 88, etc.

La psicometría ha tenido entre sus principales metas la de proporcionar al investigador herramientas conceptuales y metodológicas, que le permitan pasar de las observaciones a las mediciones. Pero, ¿cómo hacer que los resultados de la investigación psicológica se traduzcan en política y en práctica educativas?

El estudio que se expuso en esta tesis es una clara muestra de esa condición híbrida, ya que incorporó a la vez la estructura de un examen tradicional en su presentación, con los avances más recientes de la psicometría, representados por los modelos IRT. El por qué de esta circunstancia es claro en el presente caso: permitió ilustrar algunas limitaciones importantes en la evaluación, comunes en nuestro medio. Una propósito que apoyó esta decisión fue partir de una evaluación típica del desempeño escolar y utilizarla como fuente de los contrastes al hacer el reconocimiento de patrones mediante simulación neurocomputacional.

Los resultados mostraron que esto es posible, es útil y ofrece ventajas considerables para el trabajo de diagnóstico psicológico que sea susceptible de automatización. No obstante, el entorno neurocomputacional es en cierto modo ciego a las debilidades de los instrumentos de medición, porque su punto de arranque son los datos que el investigador selecciona para las fases de entrenamiento y prueba. Las debilidades de los datos se hacen patentes cuando, después de un número considerable de iteraciones (varios miles), que pueden requerir de varias horas de trabajo de máquina, un

neurocomputador no logra reducir el nivel de error y concluir exitosamente el proceso de reconocimiento, por las incongruencias en los registros.

En el presente estudio, dos rasgos sobresalientes fueron la rapidez con que se realizó el procesamiento neurocomputacional y la observación de niveles de error muy bajos, lo que aporta evidencia considerable sobre la naturalza adecuada de la información utilizada que, como fue expresado, estuvo antecedida por un análisis psicométrico extenso de los patrones y perfiles de interés.

Es importante, entonces, destacar que la calidad tanto del proceso de medición, como del de reconocimiento automatizado, dependerá en gran parte de la calidad de la información que se tome como base.

Una pregunta inquietante, porque su respuesta es difusa, se refiere a la posible correlación de la medición en una prueba de desempeño, con algunas medidas independientes pertinentes. La relación reactivo-prueba sin duda es un asunto importante, pero no dice mucho de la conexión entre la medición y la probabilidad, por ejemplo, de concluir con éxito un ciclo de estudios, de tener éxito profesional, etc., aunque así se presuponga.

Buscar y probar nexos de los resultados con indicadores relevantes a la evaluación realizada debe verse como parte integral del trabajo en curso y no como algo adicional o secundario.

En suma, como fue expresado en varios puntos de este trabajo, la labor de diagnóstico psicológico es compleja, difícil y muy demandante de habilidades profesionales altamente especializadas. Por esas razones, el beneficio que pueden aportar en su realización los desarrollos psicométricos y tecnológicos actuales pueden advertirse en este estudio, ya que constituye una muestra modesta, pero ilustrativa, de las capacidades y bondades de su aplicación al campo de la psicología.

REFERENCIAS

Andersen E.B. (1977). Sufficient statistics and latent trait models. **Psychometrika**, 42, 69-81.

Anderson J.A. (1977). Neural models with cognitive implications. En La Berge y Samuels (Eds.) **Basic Processes in reading: perception and comprehension**. Hillsdale, N.J: Erlbaum, 27-90.

Anderson J.A. (1983). Cognitive and Psychological computation with neural models. **IEEE Transactions on Systems, Man and Cybernetics**, 13, 799- 815,

Anderson J.R. (1983). **The architecture of cognition**. Cambridge, Ma: Harvard University Press.

Anderson J.R. y Bower G.H. (1973). **Human Associative Memory**. Washington D.C.: Winston.

Andrich D. (1985). A Latent Trait Model for Items with Response Dependencies: Implications for Test Construction and Analysis. En S.E. Embretson (Ed.) **Test Design. Developments in Psychology and Psychometrics**, Cap. 9. Orlando, Florida: Academic Press.

Andrich D. (1988). Rasch Models for Measurement. **Sage University Paper Series on Quantitative Applications in Social Sciences**, Num. 12152.

Barndorff-Nielsen O. (1978). **Information and Exponential Families**. New York: John Wiley and Sons.

Bransford, J.R. (1979). **Human Cognition**. Belmont, C.A.: Wadsworth.

BIGSTEPS (1994). **Rasch-Model computer program, 2.4**. MESA, University of Chicago.

Butterfield E.C., Nielsen D., Tangen K.L. y Richardson M.B. (1995). Theoretically Based Psychometric Measures of Inductive Reasoning. En S.E. Embretson (Ed.) **Test Design. Developments in Psychology and Psychometrics**, Cap. 4. Orlando, Florida: Academic Press.

California Scientific Software (1990). **The BrainMaker Users Guide and Reference Manual**. Grass Valley: CSS.

Campbell D. y Stanley J. (1970). **Diseños experimentales y cuasiexperimentales en la investigación social**. Buenos Aires: Amorrortu.

Carrillo H. (1990). Cerebro, redes neuronales y sistemas dinámicos. En **Ciencias Cognitivas: un camino hacia la mente**. ICYT, Información Científica. Abril, 163, 44-47.

Castañeda F.S. (1993). **Estructuración del conocimiento a partir de lo leído. Un marco de trabajo evaluativo**. Tesis doctoral, Facultad de Psicología, UNAM.

Castañeda F.S. y López O.M. (1988a). Nuevas perspectivas para una vieja tecnología: el texto. **Revista Interamericana de Psicología y Educación**, 1, 1, 69-82.

Castañeda F. S. y López O.M. (1988b). Estructuración de Procesos Cognoscitivos de Bajo y Alto Nivel en la Comprensión y Recuerdo de Textos de Naturaleza Científico Institucional en el Bachillerato. Número especial de la Revista **Series sobre la Universidad**, CISE, UNAM, 9, 27- 55.

Castañeda F.S. y López O.M. (1990). Inteligencia Artificial y evaluación del fracaso escolar: Un diagnosticador experto de estudiantes con riesgo. **Memorias del Seminario Internacional sobre la Implementación de la Computación en Educación**. Guatemala, International Society for Technology in Education.

Castañeda F.S. y López O.M. (1991). THOR-OMBOLO: expert system in the diagnosis of problems in text study skills in college and higher education. En M. Carretero M. Pope. R. Simons y U.I. Pozo. **Learning and Intruction**. European Research in a Internacional Context III, Oxford, Pergamon Press, 451-462.

Castañeda F.S. y López O.M. (1992). Psicología e Inteligencia Artificial. En J. Negrete (Ed.) **De la Filosofía a la Inteligencia Artificial**. México: Megabyte Noriega Editores, Limusa, Cap. 9.

Castañeda F.S., López O.M. y Ramos P.T. (1990). Evaluador computacional de estrategias cognitivas de estudio. **Memorias de la II Reunión Nacional de Pensamiento y Lenguaje**, Tuxtla Gutiérrez., Chiapas.

Castañeda, F. S., López, O.M. y Romero, J. M. (1987). Understanding the role of five induced learning strategies in science textbook comprehension. **Journal of Experimental Educación**. 55, 3, 125 -130.

Castañeda, F. S., López, O. M., Gómez, A. T., Cabrera, T. M. A. y Orozco, M.C. (1989). Evaluación metacurricular (¿desarrollo o deterioro de las habilidades de aprendizaje a partir de las prácticas docentes?). En S. Castañeda y M. López (Eds.) **La Psicología Cognoscitiva del Aprendizaje. Aprendiendo a Aprender**, México: UNAM, 147-156.

Caudill M. (1993). Neural Network Special Report. **AI Expert**. San Francisco: Miller Freeman Publication.

Caudill M. (1994). Neural Networks Primer. **AI Expert**. San Francisco: Miller Freeman Publication.

Caudill M. (1995). Using Neural Networks. **AI Expert**. San Francisco: Miller Freeman Publication.

CENEVAL (1994). Examen Nacional de Ingreso a la Educación Media Superior. Informe de Resultados 1994: Los instrumentos de evaluación. México: **Centro Nacional de Evaluación para la Educación Superior**, 3-10.

Churchland P.M. (1990), Cognitive Activity in Artificial Neural Networks. En Daniel N. Osherson y Edward Smith (Eds.). **Thinkings: An invitation to Cognitive Science**, Vol 3, 199-228. Cambridge: The MIT Press.

CISE (1995). Evaluación de habilidades cognoscitivas desde un enfoque piagetiano, en estudiantes de educación media y media superior. **Centro de Investigaciones y Servicios Educativos**. UNAM. Material inédito.

De Sánchez M. (1991). Programa de Desarrollo de Habilidades del Pensamiento (DHP). **Revista Intercontinental de Psicología y Educación**, 5, 2, 207-236.

De Vega, M. (1989). **Introducción a la Psicología Cognitiva**. Madrid, España: Alianza Psicología.

Díaz-Guerrero R. (1966). Cibernética, Psicología y Ciencias del Comportamiento. **Revista de la Universidad de México**, volumen XX, número 12, agosto. UNAM: México, páginas 11-13.

Embretson S.E. (1983). Construct validity: Construct representation versus nomothetic span. **Psychological Bulletin**, 93, 179 - 197.

Embretson S.E. (1984). A general latent trait model for response processes. **Psychometrika**, 49, 175 -186.

Embretson S.E. (1985). Introduction to the problem of test design. En S.E. Embretson (Ed.) **Test Design. Developments in Psychology and Psychometrics**. Cap. 1. Orlando, Florida: Academic Press.

Embretson S.E. (1991). A Multidimensional Latent Trait Model for Measuring Learning and Change. **Psychometrika**, 56, No. 3, 495 -515.

Espinoza I. (1990). Las redes neuronales artificiales. **Ciencias Cognitivas: un camino hacia la mente**. ICYT, Información Científica y Tecnológica. Abril, 163, 37- 43.

EXCEL (1994). Microsoft Corporation.

- Feldman J. A.* (1982) Dynamic connections in neural networks. **Biological Cybernetics**, 46, 27 - 39.
- Feldman J.A. y Ballard D.H.* (1982) Connectionist models and their properties. **Cognitive Science**, 6, 205 - 254.
- Feltovich P.J., Johnson P.E., Moller J.M. y Swanson D.B.* (1984). LCS: The role and development of medical knowledge in diagnostic expertise. En W.J. Clancey y E.H. Shortliffe (Eds.) **Readings in artificial intelligence in medicine: the first decade**. MA: Addison-Wesley.
- Gagné E.* (1985). **The cognitive psychology of school learning**. Boston, MA: Little Brown.
- Gagné R.* (1987). **Las condiciones del aprendizaje**. México: Interamericana.
- Glaser R. y Bassok M* (1989). Learning Theory and the study of Instruction. **Annual Review of Psychology**, 40, 631 -666.
- Gluck M.A. y Bower G.H.* (1988). Evaluating and Adaptive Network Model of Human Learning. **Journal of Memory and Language** 27, 166 -195.
- Grossberg S.* (1978). A theory of visual coding, memory, and development. En E.L. Leeuwenberg y J.M. Buffart (Eds.) **Formal theories of visual perception**. New York: Wiley.
- Haugeland J.* (1989). **Artificial Intelligence. The Very Idea**. Bradford Books. The MIT Press, Cambridge.
- Hawkins R. y Bower G.* (1989). **Computational Models of Learning in Simple Neural Systems**. San Diego Calif: Academic Press.
- Herrnstein, R.J., Nickerson, R.S., de Sánchez, M. Swets, J.A.* (1986). Teaching Thinking Skills. **American Psychologist**, Vol. 41, No. 11, 1279-1289.
- Hinton G.E.* (1981). Implementing semantic networks in parallel hardware. En G.E. Hinton y Anderson J.A. (Eds.) **Parallel models of associative memory**. Hillsdale: Erlbaum.
- Hinton G.E.* (1987). Connectionist learning procedures. **Technical Report**. Pittsburg: Carnegie-Mellon University.
- Hinton G.E., McClelland J.L. y Rumelhart D.E.* (1986). **Parallel distributed processing: Explorations in the microstructure of cognition, I**. Cambridge MA: Bradford Books.
- ITEMAN* (1993). **Conventional Item Analysis Program, 3.5**. Assessment Systems Corporation. St. Paul, Minnesota.

Jones P. y Hoskins J. (1987). Back-Propagation. A generalized delta learning rule. **BYTE**. October, 155-162.

Johnston, Peter (1989). **La evaluación de la comprensión lectora**. Madrid, España: Visor.

Kerlinger F. (1975). **Investigación del comportamiento (técnicas y metodología)**. 2a. edición, México: Nueva Editorial Interamericana.

Kohonen T. (1988). **Self Organization and Associative Memory** (1988). Berlin: Springer-Verlag.

Kohonen T. (1989). **Associative memory. A system-theoretic approach**. New York: Springer-Verlag.

Lawrence M. y Luedenking J. (1991). **Neural Network Computing. Introduction to Neural Networks**. California Scientific Software.

López O.M. y Castañeda F.S. (1990). ASESOR: Evaluador Electrónico de Habilidades de Estudio. **Reporte técnico**. Facultad de Psicología, UNAM.

López O.M. y Castañeda F.S. (1992). Un modelo multicomponencial y multidimensional para la evaluación integral del aprendizaje universitario. **Reporte Técnico**. Documento inédito. Universidad Nacional Autónoma de México.

López, O, Castañeda, F.S. y Gómez, A.T. (1989). Contribución a la evaluación de estrategias de aprendizaje. El inventario de habilidades de estudio (IHE). En S. Castañeda y M. López (Eds.) **La Psicología Cognositiva del Aprendizaje. Aprendiendo a Aprender**. México, UNAM, 281 -288.

López O.M., Castañeda F.S., Pineda L. y Orduña J. (1992) Nuevas perspectivas para la investigación y el diagnóstico instruccional: las redes neurales como diagnosticadoras del aprendizaje. **Revista Intercontinental de Psicología y Educación**, 5, 1, 173-194.

Lord M. F. (1980). **Applications of Item Response Theory to Practical Testing Problems**. Hillsdale New Jersey: Lawrence Earlbaum Associates. Publishers.

Massaro D.W. (1988). Some Criticisms of Connectionist Models of Human Performance. **Journal of Memory and Language**, 37, 213-234.

McClelland J.L., Rumelhart D.E. y el PDP Research Group (1986). **Parallel distributed processing: Explorations in the micorstructure of cognition** (Vol. II). Cambrige: MA: Bradford Books.

McClelland J.L. (1988). Connectionist Models and Psychological Evidence. **Journal of Memory and Language**, 27, 107-123.

McCulloch W.S. y Pitts W.H. (1943). A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematic Biophysics**, 5, 115-133.

Minsky M. y Paper S. (1969), **Perceptrons: An introduction to computational geometry**. Cambridge, MA: MIT Press.

Negrete, J. (1966). Cibernética y Aspectos Dinámicos en Biología. **Revista de la Universidad de México**, volumen XX, número 12, agosto. UNAM: México, páginas 30-32.

Negrete M. J. (1988). **Inteligencia Experimental en Computadoras**. México: Limusa.

Negrete M. J. (1990). **Inteligencia, aunque sea artificial**. México: Limusa-Noriega.

Negrete M.J., Castañeda F.S y López O.M. (1992). Hacia una formalización de la conducta del perito. En J. Negrete (Ed.) **De la Filosofía a la Inteligencia Artificial**, capítulo 9. México: Megabyte, Noriega Editores, Limusa.

Neural Works (1989). **Neural Works Professional II**. Neural Ware Inc.

Pellegrino, J.W., Mumaw R.J. y Shute V.J. (1985). Analysis of spatial aptitude and expertise. En S.E. Embretson (Ed.) **Test Design. Developments in Psychology and Psychometrics**. Cap. 3. Orlando, Florida: Academic Press.

Pozo, J. I. (1990). **Teorías cognitivas del aprendizaje**. Madrid, España: Morata.

RASCAL (1992). **Rasch Analysis Program, 3.5**. Assessment Systems Corporation. St. Paul, Minnesota.

Rasch, C (1960) **Probabilistic Models for some Intelligence and Attainment Tests**. Copenhagen: Danish Institute for Educational Research, 1960. (Expanded edition, Chicago: The University of Chicago press.

Reyes Lagunes I (1992). Evaluación Educativa: Una revisión. **Revista Intercontinental de Psicología y Educación**, 5, 2, 195-206.

Reyes D. y Piña J. (1980). **De la primaria a la secundaria**. México: Ediciones Pedagógicas.

Rojas E. y Garda A. Inteligencia Artificial y Ciencia Cognitiva en México (1980). **Ciencias Cognitivas: un camino hacia la mente**. ICYT, Información Científica y Tecnológica. Abril, 163, 57-60.

Rosenberg C.H. y Blellock G. (1988). **Connectionist Models. An implementation of Network Learning on the Connection Machine**, Cap. 12. MIT AI Laboratory. Princeton University.

Rumelhart D.E., Hinton G.E. y E.J. Williams (1986). Learning Internal Representations by Error Propagation. En **Parallel distributed processing: Explorations in the microstructure of cognition**, I. Cambridge MA: Bradford Books.

SEP (1994). **Plan y Programas de Estudio. Educación Básica**. Dirección General de Materiales y Métodos Educativos de la Subsecretaría de Educación Básica y Normal, segunda edición. México, D.F.: Secretaría de Educación Pública.

Sejnowski T.J. y Rosenberg C.R. (1986). NET-Talk: A Parallel Network That Learns to Read Aloud. **Neural Computing, Foundations of Research**. Cambridge: MIR Press.

Shiple Ch. (1989). Whatever Happened to AI? **PC Computing**. March, 64-74.

Simon H.A. (1979). **Las Ciencias de lo Artificial**. España: ATE.

Sklansky J. y Wassel G.N. (1981). **Pattern Classifiers and Trainable Machines**. New York: Springer-Verlag.

Snow R. y Swanson J. (1985). Instructional Psychology: Aptitude, Adaptation, and Assessment School of Education. **California Rev. Psychology**, 43, 538 - 626.

Sternberg R.J. (1977). **Intelligence, information processing, and analogical reasoning: the componential analysis of human abilities**. Hillsdale N.J: Erlbaum.

Sternberg R.J. y Mc Namara T.P. (1985). The representation and Processing of Information in real Time Verbal Comprehension. En S.E. Embretson (Ed.) **Test Design. Developments in Psychology and Psychometrics**. Cap. 2. Orlando, Florida: Academic Press.

SPSS (1993). **SPSS. Base System Syntax**. Reference Guide, Release 6.0. Chicago, Ill.

SYSTAT (1992). **Systat Inc. Intelligent Software**, 5.0. Evanstone, Ill.

Thorndike, R. y Lohman D.A. (1990). **A Century of Ability Testing**. The Riverside Publishing Company.

Thorndike, R. (1989) **Psicometría Aplicada**. México: Limusa.

Waltz, D.L. y Pollack, J. (1986) Massively Parallel Parsing: A strongly Interactive Model of Natural Language Interpretation. En D. Waltz y J.A.

Feldman (Eds.) **Connectionist Model and their Implications: Readings from Cognitive Science**. Ablex Publishing Co.

Wildrow G. y M.E. Hoff (1960). Adaptive Switching Circuits. Institute of Radio Engineers, **Convention Record**.

Widrow B. (1963/1987). Adaline and Madelina. **First International Conference on Neural Networks**, Vol. II. Piscataway: IEEE.

White H. (1989). Neural-Networks Learning and Statistics. **AI Expert**, Dec., 48 - 52.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. **Journal of Educational Measurement**, 14, 97 - 116.

Wright B. y Masters G. (1982). **Rating Scale Analysis**. Mesa Press. University of Chicago.

Wright B. y Stone H. (1979). **Best Test Desing**. Mesa Press. University of Chicago.

Woodrow H. (1938). The relationship between abilities and improvement with practice. **Journal of Educational Psychology**, 29, 215-230. [No incluir.

XCALIBRE (1995). **Marginal Maximum-Likelihood Estimation Program, 1.0**. Assessment Systems Corporation. St. Paul, Minnesota.

ANEXOS

1. Resultados de los 300 sujetos del grupo 1

2. Resultados en cada reactivo

Anexo 1: Resultados de los 300 sujetos del grupo 1, expresados en total de aciertos, proporción de respuestas correctas en las subpruebas, habilidades y tipos de conocimiento, diferenciando los subgrupos por nivel de desempeño. Se indica también en *cursivas* el total de reactivos correspondiente a cada indicador. La *n* de cada subgrupo está dada por los rangos en el total de aciertos. Los sujetos (SS) están ordenados en forma decreciente, según su puntuación.

A. DESEMPEÑO ALTO

n = 96			HABILIDADES			CONOCIMIENT.			MEDICIÓN EN LAS SUBPRUEBAS							
SS	SUM	PROP	IDE	INF	RES	CON	FAC	PRO	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO
			<i>58</i>	<i>29</i>	<i>33</i>	<i>77</i>	<i>22</i>	<i>21</i>	<i>24</i>	<i>12</i>	<i>12</i>	<i>12</i>	<i>24</i>	<i>12</i>	<i>12</i>	<i>12</i>
1	116	0.97	0.97	0.97	0.97	0.97	0.91	1.00	0.92	0.83	1.00	1.00	1.00	1.00	1.00	1.00
2	115	0.96	0.98	0.90	0.97	0.94	1.00	1.00	0.96	0.92	1.00	1.00	1.00	0.92	1.00	0.83
3	115	0.96	0.95	0.93	1.00	0.95	0.95	1.00	0.92	0.92	0.92	1.00	0.96	1.00	1.00	1.00
4	114	0.95	0.98	0.86	0.97	0.95	0.95	0.95	0.88	0.92	1.00	0.92	0.96	1.00	1.00	1.00
5	114	0.95	0.95	0.90	1.00	0.96	0.86	1.00	0.92	0.92	0.92	1.00	1.00	1.00	0.92	0.92
6	112	0.93	0.97	0.79	1.00	0.92	0.91	1.00	0.83	1.00	1.00	0.92	1.00	0.92	1.00	0.83
7	111	0.93	0.95	0.83	0.97	0.92	0.91	0.95	0.79	0.92	1.00	1.00	0.98	1.00	0.92	0.92
8	110	0.92	0.97	0.83	0.91	0.91	0.91	0.95	0.88	0.83	0.83	1.00	0.92	1.00	1.00	0.92
9	109	0.91	0.95	0.79	0.94	0.88	0.95	0.95	0.88	0.75	0.92	1.00	0.92	0.92	1.00	0.92
10	108	0.90	0.86	0.90	0.97	0.91	0.82	0.95	0.83	0.75	0.92	1.00	0.96	1.00	0.92	0.83
11	108	0.90	0.90	0.79	1.00	0.90	0.82	1.00	0.83	0.83	0.83	1.00	1.00	1.00	0.83	0.83
12	108	0.90	0.90	0.90	0.91	0.91	0.82	0.95	0.92	0.83	0.92	1.00	0.83	1.00	0.83	0.92
13	105	0.88	0.90	0.76	0.94	0.87	0.77	1.00	0.83	0.67	0.92	0.92	0.96	0.92	0.92	0.83
14	104	0.87	0.86	0.86	0.88	0.90	0.82	0.81	0.79	0.92	0.92	0.92	0.88	0.92	0.92	0.75
15	103	0.86	0.88	0.76	0.91	0.82	0.95	0.90	0.71	1.00	0.92	0.75	0.92	0.92	0.92	0.83
16	102	0.85	0.86	0.86	0.82	0.88	0.82	0.76	0.83	0.92	0.83	0.92	0.83	0.67	1.00	0.83
17	102	0.85	0.88	0.79	0.85	0.90	0.77	0.76	0.88	0.92	0.75	0.83	0.83	0.92	0.75	0.92
18	101	0.84	0.84	0.90	0.79	0.84	0.86	0.81	0.83	1.00	0.92	0.83	0.83	0.75	0.83	0.75
19	101	0.84	0.84	0.86	0.82	0.86	0.77	0.86	0.83	1.00	0.67	0.83	0.88	0.67	1.00	0.83
20	100	0.83	0.88	0.72	0.85	0.81	0.95	0.81	0.75	0.67	1.00	0.83	0.83	0.92	0.92	0.83
21	99	0.83	0.88	0.79	0.76	0.83	0.86	0.76	0.83	0.92	1.00	0.83	0.67	0.83	0.83	0.83
22	98	0.82	0.81	0.76	0.88	0.82	0.68	0.95	0.79	0.67	0.58	0.92	0.92	0.83	1.00	0.75
23	98	0.82	0.79	0.72	0.94	0.77	0.95	0.86	0.67	0.92	1.00	0.67	0.83	0.83	0.92	0.83
24	98	0.82	0.81	0.93	0.73	0.83	0.86	0.71	0.83	0.92	0.92	0.92	0.71	0.83	0.75	0.75
25	98	0.82	0.88	0.69	0.82	0.81	0.86	0.81	0.63	0.92	0.83	0.83	0.75	0.92	1.00	0.92
26	97	0.81	0.84	0.69	0.85	0.78	0.86	0.86	0.75	0.83	0.92	0.83	0.79	0.83	0.67	0.92
27	97	0.81	0.79	0.72	0.91	0.83	0.64	0.90	0.75	0.67	0.75	1.00	0.96	0.92	0.58	0.75
28	96	0.80	0.83	0.66	0.88	0.79	0.77	0.86	0.58	0.83	0.75	0.92	0.83	0.75	1.00	0.92
29	95	0.79	0.81	0.72	0.82	0.78	0.82	0.81	0.67	0.75	0.92	0.83	0.88	0.67	0.83	0.83

SS	SUM	PROP	IDE	INF	RES	CON	FAC	PRO	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO
187	43	0.36	0.41	0.17	0.42	0.35	0.32	0.43	0.21	0.42	0.33	0.33	0.50	0.33	0.33	0.42
188	43	0.36	0.34	0.34	0.39	0.34	0.32	0.48	0.33	0.42	0.42	0.25	0.33	0.67	0.33	0.17
189	43	0.36	0.45	0.28	0.27	0.36	0.41	0.29	0.29	0.17	0.58	0.33	0.42	0.42	0.33	0.33
190	43	0.36	0.41	0.28	0.33	0.36	0.41	0.29	0.29	0.33	0.33	0.50	0.25	0.50	0.58	0.25
191	43	0.36	0.36	0.28	0.42	0.32	0.36	0.48	0.33	0.25	0.17	0.33	0.21	0.42	0.67	0.67
192	43	0.36	0.34	0.34	0.39	0.39	0.14	0.48	0.46	0.17	0.17	0.25	0.50	0.33	0.58	0.17
193	43	0.36	0.33	0.38	0.39	0.35	0.36	0.38	0.29	0.58	0.17	0.42	0.33	0.50	0.42	0.25
194	43	0.36	0.38	0.41	0.27	0.40	0.36	0.19	0.29	0.67	0.50	0.33	0.13	0.42	0.17	0.67
195	42	0.35	0.43	0.28	0.27	0.40	0.32	0.19	0.29	0.33	0.25	0.42	0.38	0.42	0.33	0.42
196	42	0.35	0.36	0.31	0.36	0.36	0.32	0.33	0.29	0.25	0.42	0.25	0.38	0.42	0.58	0.25
197	42	0.35	0.29	0.31	0.48	0.26	0.50	0.52	0.25	0.25	0.17	0.50	0.46	0.50	0.25	0.42
198	42	0.35	0.38	0.38	0.27	0.43	0.18	0.24	0.42	0.25	0.42	0.33	0.17	0.42	0.50	0.42
199	42	0.35	0.34	0.34	0.36	0.38	0.23	0.38	0.38	0.33	0.42	0.33	0.29	0.33	0.42	0.33
200	42	0.35	0.43	0.21	0.33	0.38	0.41	0.19	0.29	0.33	0.33	0.50	0.29	0.50	0.33	0.33
201	41	0.34	0.31	0.34	0.39	0.32	0.32	0.43	0.29	0.42	0.25	0.25	0.33	0.50	0.50	0.25
202	41	0.34	0.34	0.31	0.36	0.36	0.32	0.29	0.42	0.33	0.17	0.17	0.33	0.50	0.50	0.25
203	41	0.34	0.28	0.41	0.39	0.31	0.36	0.43	0.29	0.33	0.25	0.42	0.38	0.33	0.33	0.42

C. DESEMPEÑO BAJO

n = 97

			HABILIDADES			CONOCIMIENT.			MEDICIÓN EN LAS SUBPRUEBAS							
SS	SUM	PROP	IDE	INF	RES	CON	FAC	PRO	ESP	LIT	HIS	GEO	MAT	FIS	QUI	BIO
			58	29	33	77	22	21	24	12	12	12	24	12	12	12
204	40	0.33	0.33	0.31	0.36	0.34	0.27	0.38	0.21	0.42	0.08	0.25	0.42	0.25	0.67	0.42
205	40	0.33	0.43	0.28	0.21	0.36	0.41	0.14	0.33	0.50	0.25	0.58	0.21	0.33	0.17	0.42
206	40	0.33	0.31	0.34	0.36	0.32	0.32	0.38	0.21	0.33	0.17	0.33	0.33	0.42	0.50	0.50
207	40	0.33	0.38	0.31	0.27	0.36	0.32	0.24	0.29	0.42	0.25	0.58	0.21	0.25	0.42	0.42
208	40	0.33	0.31	0.24	0.45	0.34	0.23	0.43	0.29	0.17	0.42	0.08	0.38	0.42	0.25	0.67
209	39	0.33	0.33	0.45	0.21	0.31	0.55	0.14	0.46	0.42	0.25	0.50	0.25	0.25	0.25	0.17
210	39	0.33	0.29	0.41	0.30	0.34	0.36	0.24	0.46	0.25	0.58	0.25	0.29	0.25	0.00	0.42
211	39	0.33	0.34	0.24	0.36	0.30	0.36	0.38	0.17	0.33	0.33	0.42	0.25	0.50	0.50	0.33
212	39	0.33	0.31	0.41	0.27	0.34	0.27	0.33	0.42	0.50	0.25	0.25	0.17	0.33	0.58	0.17
213	39	0.33	0.34	0.24	0.36	0.29	0.32	0.48	0.21	0.08	0.50	0.33	0.46	0.33	0.50	0.17
214	39	0.33	0.33	0.48	0.18	0.30	0.50	0.24	0.38	0.50	0.50	0.50	0.13	0.17	0.42	0.17
215	39	0.33	0.41	0.17	0.30	0.34	0.27	0.33	0.17	0.17	0.33	0.58	0.21	0.42	0.42	0.58
216	39	0.33	0.33	0.38	0.27	0.34	0.36	0.24	0.33	0.42	0.33	0.25	0.33	0.25	0.33	0.33
217	39	0.33	0.38	0.24	0.30	0.36	0.27	0.24	0.29	0.25	0.08	0.25	0.33	0.33	0.25	0.83
218	39	0.33	0.28	0.38	0.36	0.35	0.27	0.29	0.29	0.42	0.17	0.25	0.33	0.42	0.67	0.08
219	38	0.32	0.29	0.31	0.36	0.29	0.32	0.43	0.17	0.33	0.42	0.42	0.38	0.25	0.42	0.25

Anexo 2: Resultados en cada reactivo, indicando la asignatura, habilidad y tipo de conocimiento que mide y la proporción de aciertos obtenidos.

Tabla 1. Número de aciertos en cada reactivo.

Ítem	Asign.	Habilid.	Conoc.	Aclert.
1	ESP	INF	CON	0.65
2	ESP	INF	CON	0.44
3	ESP	INF	CON	0.5
4	ESP	INF	CON	0.44
5	ESP	INF	CON	0.52
6	ESP	INF	CON	0.61
7	ESP	INF	CON	0.35
8	ESP	INF	CON	0.5
9	ESP	INF	CON	0.42
10	ESP	INF	CON	0.43
11	ESP	INF	CON	0.4
12	ESP	RES	CON	0.37
13	ESP	IDE	CON	0.67
14	ESP	IDE	CON	0.34
15	ESP	RES	CON	0.57
16	ESP	IDE	CON	0.59
17	ESP	IDE	CON	0.25
18	ESP	IDE	CON	0.75
19	ESP	INF	CON	0.32
20	ESP	IDE	CON	0.55

Ítem	Asign.	Habilid.	Conoc.	Aclert.
21	ESP	INF	CON	0.6
22	ESP	INF	CON	0.5
23	ESP	INF	CON	0.27
24	ESP	INF	CON	0.1
25	LIT	IDE	CON	0.52
26	LIT	RES	CON	0.56
27	LIT	RES	FAC	0.44
28	LIT	IDE	FAC	0.62
29	LIT	IDE	FAC	0.5
30	LIT	IDE	CON	0.62
31	LIT	IDE	CON	0.55
32	LIT	IDE	FAC	0.52
33	LIT	IDE	CON	0.62
34	LIT	INF	CON	0.47
35	LIT	INF	CON	0.57
36	LIT	IDE	FAC	0.38
37	HIS	INF	FAC	0.5
38	HIS	INF	CON	0.54
39	HIS	IDE	FAC	0.53
40	HIS	IDE	CON	0.42

Ítem	Asign.	Habilid.	Conoc.	Aclert.
41	HIS	IDE	FAC	0.29
42	HIS	INF	CON	0.51
43	HIS	IDE	FAC	0.43
44	HIS	IDE	FAC	0.57
45	HIS	INF	CON	0.49
46	HIS	IDE	FAC	0.59
47	HIS	IDE	CON	0.55
48	HIS	IDE	FAC	0.56
49	GEO	IDE	CON	0.59
50	GEO	IDE	CON	0.42

Ítem	Asign.	Habilid.	Conoc.	Aclert.
51	GEO	IDE	CON	0.49
52	GEO	IDE	FAC	0.56
53	GEO	IDE	FAC	0.46
54	GEO	IDE	CON	0.46
55	GEO	IDE	CON	0.64
56	GEO	IDE	CON	0.36
57	GEO	INF	CON	0.61
58	GEO	INF	FAC	0.43
59	GEO	IDE	FAC	0.48
60	GEO	IDE	FAC	0.55

Item	Asign.	Habilld.	Conoc.	Aclert.
61	MAT	RES	CON	0.67
62	MAT	IDE	CON	0.52
63	MAT	IDE	CON	0.35
64	MAT	RES	PRO	0.46
65	MAT	RES	CON	0.51
66	MAT	RES	PRO	0.37
67	MAT	IDE	CON	0.45
68	MAT	IDE	PRO	0.46
69	MAT	RES	PRO	0.48
70	MAT	RES	PRO	0.45

Item	Asign.	Habilld.	Conoc.	Aclert.
71	MAT	RES	PRO	0.57
72	MAT	RES	PRO	0.42
73	MAT	RES	PRO	0.43
74	MAT	RES	PRO	0.42
75	MAT	RES	CON	0.62
76	MAT	RES	PRO	0.44
77	MAT	RES	PRO	0.52
78	MAT	RES	PRO	0.44
79	MAT	IDE	CON	0.43
80	MAT	RES	CON	0.55

Item	Asign.	Habilld.	Conoc.	Aclert.
81	MAT	RES	CON	0.39
82	MAT	RES	PRO	0.46
83	MAT	IDE	CON	0.52
84	MAT	RES	PRO	0.57
85	FIS	RES	PRO	0.6
86	FIS	IDE	CON	0.53
87	FIS	RES	PRO	0.49
88	FIS	INF	FAC	0.56
89	FIS	RES	CON	0.6
90	FIS	IDE	FAC	0.47
91	FIS	RES	CON	0.44
92	FIS	RES	PRO	0.5
93	FIS	RES	PRO	0.39
94	FIS	IDE	CON	0.51
95	FIS	RES	PRO	0.53
96	FIS	RES	CON	0.57
97	QUI	IDE	CON	0.46
98	QUI	IDE	CON	0.51
99	QUI	IDE	CON	0.69
100	QUI	IDE	CON	0.45

Item	Asign.	Habilld.	Conoc.	Aclert.
101	QUI	IDE	CON	0.48
102	QUI	IDE	CON	0.64
103	QUI	INF	CON	0.34
104	QUI	IDE	CON	0.73
105	QUI	RES	PRO	0.47
106	QUI	INF	FAC	0.56
107	QUI	IDE	CON	0.47
108	QUI	RES	PRO	0.74
109	BIO	IDE	FAC	0.53
110	BIO	IDE	CON	0.49
111	BIO	RES	CON	0.64
112	BIO	IDE	FAC	0.42
113	BIO	IDE	CON	0.63
114	BIO	IDE	CON	0.52
115	BIO	IDE	CON	0.49
116	BIO	INF	CON	0.2
117	BIO	IDE	CON	0.55
118	BIO	IDE	CON	0.5
119	BIO	IDE	CON	0.61
120	BIO	INF	CON	0.55

La tabla 2 presentará los resultados del análisis de los reactivos, en la *secuencia en que fueron administrados* (1 a 120) en las ocho subpruebas. Las primeras dos columnas de la tabla identifican el *número de reactivo* y su *contenido*. La tercera, *proporción de aciertos*, informa ese resultado en cada ítem y permite distinguir los valores cercanos a cero (muy difíciles) y a uno (muy fáciles).

En la tabla se indica el contenido que midió la subprueba: BIO (biología), ESP (español), FIS (física), GEO (geografía), HIS (historia), LIT (literatura), MAT (matemáticas) y QUI (química).

La columna que muestra el índice de discriminación expresa la capacidad de cada reactivo para diferenciar los sujetos con *alto o bajo nivel de desempeño*, por la diferencia entre $P_{\text{alto}} - P_{\text{bajo}}$, donde P_{alto} es la proporción de examinados arriba del 27 % de la distribución y P_{bajo} es la misma proporción del grupo bajo en el número de aciertos. Para prestar atención a los aspectos más globales de la información, se omitió en esta tabla el comportamiento de las opciones de cada ítem.

Dada la forma dicotómica de los datos, se consideró que las formas de análisis de *correlación biserial* -de uso más común- y *punto biserial* -más estable entre muestras- son aplicables para determinar índices de discriminación en los reactivos, por lo que en la tabla se ofrecen ambas variantes como indicadores alternativos. Son correlaciones *reactivo-total*, es decir, operan entre los puntajes correctos y el número total de aciertos en la prueba. Aunque en la tabla se advierte que los valores de ambos coeficientes de correlación difieren entre sí, muestran las mismas tendencias. Como podrá advertirse, los reactivos 24 y 116 mostraron valores negativos en la discriminación, indicando claramente que son defectuosos.

Tabla 2. Análisis psicométrico de los reactivos.

Orden de presentación en el examen

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl. Biser.	Point Biser.
1	ESP	0.65	0.47	0.52	0.40
2	ESP	0.44	0.44	0.48	0.38
3	ESP	0.5	0.64	0.72	0.57
4	ESP	0.44	0.57	0.66	0.53
5	ESP	0.52	0.61	0.63	0.50
6	ESP	0.61	0.19	0.26	0.20
7	ESP	0.35	0.50	0.56	0.43
8	ESP	0.5	0.72	0.75	0.60
9	ESP	0.42	0.72	0.80	0.64
10	ESP	0.43	0.66	0.74	0.59
11	ESP	0.4	0.58	0.64	0.50
12	ESP	0.37	0.67	0.73	0.57

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl. Biser.	Point Biser.
13	ESP	0.67	0.57	0.62	0.48
14	ESP	0.34	0.48	0.57	0.44
15	ESP	0.57	0.65	0.67	0.53
16	ESP	0.59	0.46	0.48	0.38
17	ESP	0.25	0.28	0.31	0.23
18	ESP	0.75	0.45	0.54	0.40
19	ESP	0.32	0.49	0.56	0.43
20	ESP	0.55	0.48	0.52	0.42
21	ESP	0.6	0.65	0.62	0.49
22	ESP	0.5	0.62	0.61	0.49
23	ESP	0.27	0.18	0.23	.017
24	ESP	0.10	-0.13	-0.26	-0.15

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
25	LIT	0.52	0.53	0.54	0.43
26	LIT	0.56	0.48	0.50	0.40
27	LIT	0.44	0.54	0.63	0.50
28	LIT	0.62	0.43	0.48	0.37
29	LIT	0.50	0.63	0.63	0.50
30	LIT	0.62	0.54	0.62	0.48
31	LIT	0.55	0.66	0.70	0.55
32	LIT	0.52	0.36	0.32	0.26
33	LIT	0.62	0.45	0.49	0.38
34	LIT	0.47	0.48	0.47	0.38
35	LIT	0.57	0.57	0.63	0.50
36	LIT	0.38	0.38	0.47	0.37

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
37	HIS	0.50	0.38	0.39	0.31
38	HIS	0.54	0.62	0.67	0.53
39	HIS	0.53	0.62	0.64	0.51
40	HIS	0.42	0.41	0.47	0.37
41	HIS	0.29	0.47	0.55	0.42
42	HIS	0.51	0.67	0.65	0.51
43	HIS	0.43	0.56	0.59	0.47
44	HIS	0.57	0.5	0.53	0.42
45	HIS	0.49	0.49	0.44	0.35
46	HIS	0.59	0.65	0.67	0.53
47	HIS	0.55	0.67	0.74	0.59
48	HIS	0.56	0.60	0.64	0.51

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
49	GEO	0.59	0.32	0.35	0.28
50	GEO	0.42	0.54	0.50	0.39
51	GEO	0.49	0.52	0.53	0.42
52	GEO	0.56	0.52	0.54	0.43
53	GEO	0.46	0.59	0.66	0.53
54	GEO	0.46	0.42	0.47	0.38
55	GEO	0.64	0.37	0.47	0.37
56	GEO	0.36	0.51	0.60	0.47
57	GEO	0.61	0.66	0.68	0.53
58	GEO	0.43	0.59	0.58	0.46
59	GEO	0.48	0.59	0.59	0.47
60	GEO	0.55	0.74	0.73	0.58

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
61	MAT	0.67	0.57	0.62	0.47
62	MAT	0.52	0.53	0.55	0.44
63	MAT	0.35	0.50	0.56	0.43
64	MAT	0.46	0.37	0.42	0.34
65	MAT	0.51	0.54	0.61	0.49
66	MAT	0.37	0.50	0.53	0.41
67	MAT	0.45	0.51	0.55	0.44
68	MAT	0.46	0.54	0.60	0.48
69	MAT	0.48	0.62	0.60	0.48
70	MAT	0.45	0.54	0.58	0.46
71	MAT	0.57	0.59	0.63	0.50
72	MAT	0.42	0.42	0.51	0.40

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
73	MAT	0.43	0.46	0.50	0.40
74	MAT	0.42	0.50	0.48	0.38
75	MAT	0.62	0.45	0.51	0.40
76	MAT	0.44	0.51	0.53	0.42
77	MAT	0.52	0.54	0.59	0.47
78	MAT	0.44	0.57	0.60	0.47
79	MAT	0.43	0.48	0.50	0.40
80	MAT	0.55	0.66	0.70	0.56
81	MAT	0.39	0.60	0.64	0.50
82	MAT	0.46	0.50	0.53	0.43
83	MAT	0.52	0.49	0.56	0.45
84	MAT	0.57	0.59	0.63	0.50

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
85	FÍS	0.60	0.64	0.64	0.51
86	FÍS	0.53	0.42	0.45	0.36
87	FÍS	0.49	0.59	0.58	0.47
88	FÍS	0.56	0.56	0.56	0.44
89	FÍS	0.60	0.51	0.54	0.42
90	FÍS	0.47	0.42	0.50	0.40
91	FÍS	0.44	0.58	0.59	0.47
92	FÍS	0.50	0.71	0.73	0.59
93	FÍS	0.39	0.52	0.52	0.41
94	FÍS	0.51	0.65	0.65	0.52
95	FÍS	0.53	0.38	0.39	0.31
96	FÍS	0.57	0.52	0.50	0.40

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
97	QUI	0.46	0.50	0.55	0.44
98	QUI	0.51	0.58	0.59	0.47
99	QUI	0.69	0.44	0.52	0.40
100	QUI	0.45	0.42	0.48	0.39
101	QUI	0.48	0.24	0.30	0.24
102	QUI	0.64	0.52	0.57	0.45
103	QUI	0.34	0.52	0.64	0.49
104	QUI	0.73	0.23	0.32	0.24
105	QUI	0.47	0.46	0.52	0.41
106	QUI	0.56	0.56	0.60	0.48
107	QUI	0.47	0.47	0.51	0.41
108	QUI	0.74	0.46	0.57	0.42

Sec.	Asign.	Prop. Acier.	Índice Discrim.	Correl Biser.	Point Biser
109	BIO	0.53	0.62	0.65	0.52
110	BIO	0.49	0.61	0.67	0.54
111	BIO	0.64	0.72	0.72	0.56
112	BIO	0.42	0.30	0.32	0.25
113	BIO	0.63	0.54	0.61	0.48
114	BIO	0.52	0.71	0.71	0.57
115	BIO	0.49	0.79	0.80	0.64
116	BIO	0.20	-0.05	-0.04	-0.03
117	BIO	0.55	0.52	0.54	0.43
118	BIO	0.50	0.60	0.61	0.49
119	BIO	0.61	0.60	0.61	0.48
120	BIO	0.55	0.38	0.37	0.30

Para favorecer el análisis de los resultados centrados en las características de la prueba, se presentarán los datos separando los reactivos en los tres grados de dificultad ya mencionados (fácil, medio y difícil) e indicando su nivel de dificultad.

Tabla 3. Orden de dificultad de los reactivos, de menor a mayor.

En la tabla 3 están enmarcados los diez reactivos más fáciles y los más difíciles, para hacer un análisis particular sobre sus características psicométricas. El corte en los niveles está dado por el número de aciertos, tomando como referencia los puntajes crudos cada 40 reactivos. La tabla permite advertir que 62 reactivos tuvieron una proporción de aciertos entre 0.10 y 0.50, lo que constituye el 52 % y los 58 reactivos restantes estuvieron entre 0.51 y 0.75.

Sujetos del grupo 1 (n = 300)

A. DIFICULTAD BAJA

Ítem	Asign.	Habilid	Conoc.	Dificul.	Punt.	Prop.	Orden
18	Esp	IDE	CON	75	228	0.76	1
104	Quí	IDE	CON	71	222	0.74	2
108	Quí	RES	PRO	68	222	0.74	3
99	Quí	IDE	CON	65	208	0.69	4
5	Esp	INF	CON	49	203	0.68	5
61	Mat	RES	CON	59	203	0.68	6
13	Esp	IDE	CON	61	199	0.66	7
1	Esp	INF	CON	62	196	0.65	8
111	Bio	RES	CON	63	195	0.65	9
55	Geo	IDE	CON	59	194	0.65	10
102	Quí	IDE	CON	61	194	0.65	11
113	Bio	IDE	CON	58	190	0.63	12
30	Lit	IDE	CON	57	188	0.63	13
33	Lit	IDE	CON	61	186	0.62	14
75	Mat	RES	CON	49	186	0.62	15
6	Esp	INF	CON	64	185	0.62	16
28	Lit	IDE	FAC	56	185	0.62	17
57	Geo	INF	CON	57	184	0.61	18
119	Bio	IDE	CON	53	184	0.61	19
85	Fís	RES	PRO	55	182	0.61	20
89	Fís	RES	CON	51	181	0.60	21
21	Esp	INF	CON	57	178	0.59	22
46	His	IDE	FAC	55	178	0.59	23
49	Geo	IDE	CON	59	178	0.59	24
16	Esp	IDE	CON	57	177	0.59	25

Ítem	Asign.	Habilid	Conoc.	Dificul.	Punt.	Prop.	Orden
71	Mat	RES	PRO	52	173	0.58	26
84	Mat	RES	PRO	47	172	0.57	27
35	Lit	INF	CON	51	171	0.57	28
96	Fís	RES	CON	52	171	0.57	29
44	His	IDE	FAC	48	170	0.57	30
15	Esp	RES	CON	49	169	0.56	31
52	Geo	IDE	FAC	54	169	0.56	32
88	Fís	INF	FAC	53	169	0.56	33
26	Lit	RES	CON	52	168	0.56	34
106	Quí	INF	FAC	55	168	0.56	35
60	Geo	IDE	FAC	53	167	0.56	36
48	His	IDE	FAC	53	166	0.55	37
80	Mat	RES	CON	48	166	0.55	38
117	Bio	IDE	CON	48	166	0.55	39
120	Bio	INF	CON	53	166	0.55	40
Media				56.25	183.93	0.61	

B. DIFICULTAD MEDIA

Ítem	Asign.	Habilid	Conoc.	Dificul.	Punt.	Prop.	Orden
20	Esp	IDE	CON	52	165	0.55	41
47	His	IDE	CON	51	165	0.55	42
38	His	INF	CON	49	164	0.55	43
31	Lit	IDE	CON	51	163	0.54	44
95	Fís	RES	PRO	54	162	0.54	45
39	His	IDE	FAC	44	159	0.53	46
86	Fís	IDE	CON	49	159	0.53	47
25	Lit	IDE	CON	48	158	0.53	48
109	Bio	IDE	FAC	50	158	0.53	49
32	Lit	IDE	FAC	51	156	0.52	50
62	Mat	IDE	CON	46	156	0.52	51
77	Mat	RES	PRO	49	156	0.52	52
83	Mat	IDE	CON	45	156	0.52	53
114	Bio	IDE	CON	43	156	0.52	54
42	His	INF	CON	49	155	0.52	55
98	Quí	IDE	CON	46	155	0.52	56

Ítem	Asign.	Habilid	Conoc.	Dificul.	Punt.	Prop.	Orden
65	Mat	RES	CON	47	154	0.51	57
94	Fís	IDE	CON	45	154	0.51	58
118	Bio	IDE	CON	47	153	0.51	59
22	Esp	INF	CON	42	152	0.51	60
92	Fís	RES	PRO	44	151	0.50	61
3	Esp	INF	CON	42	150	0.50	62
8	Esp	INF	CON	44	150	0.50	63
29	Lit	IDE	FAC	45	150	0.50	64
37	His	INF	FAC	49	150	0.50	65
110	Bio	IDE	CON	43	149	0.50	66
87	Fís	RES	PRO	43	148	0.49	67
45	His	INF	CON	43	147	0.49	68
51	Geo	IDE	CON	41	147	0.49	69
115	Bio	IDE	CON	44	147	0.49	70
101	Quí	IDE	CON	49	146	0.49	71
69	Mat	RES	PRO	39	145	0.48	72
59	Geo	IDE	FAC	44	144	0.48	73
90	Fís	IDE	FAC	44	142	0.47	74
107	Quí	IDE	CON	38	142	0.47	75
105	Quí	RES	PRO	40	141	0.47	76
53	Geo	IDE	FAC	40	140	0.47	77
54	Geo	IDE	CON	41	140	0.47	78
97	Quí	IDE	CON	46	140	0.47	79
34	Lit	INF	CON	49	139	0.46	80
Media				45.65	151.60	0.51	

C. DIFICULTAD ALTA

Ítem	Asign.	Habilid	Conoc.	Dificul.	Punt.	Prop.	Orden
68	Mat	IDE	PRO	35	138	0.46	81
64	Mat	RES	PRO	37	137	0.46	82
82	Mat	RES	PRO	41	137	0.46	83
67	Mat	IDE	CON	43	136	0.45	84
100	Quí	IDE	CON	35	135	0.45	85
27	Lit	RES	FAC	35	134	0.45	86

Item	Asign.	Habllid	Conoc.	Dificul.	Punt.	Prop.	Orden
70	Mat	RES	PRO	37	134	0.45	87
4	Esp	INF	CON	37	133	0.44	88
78	Mat	RES	PRO	39	133	0.44	89
91	Fis	RES	CON	35	133	0.44	90
2	Esp	INF	CON	38	132	0.44	91
58	Geo	INF	FAC	37	131	0.44	92
76	Mat	RES	PRO	33	131	0.44	93
43	His	IDE	FAC	40	130	0.43	94
73	Mat	RES	PRO	40	130	0.43	95
79	Mat	IDE	CON	38	130	0.43	96
112	Bio	IDE	FAC	40	128	0.43	97
10	Esp	INF	CON	42	127	0.42	98
72	Mat	RES	PRO	33	126	0.42	99
9	Esp	INF	CON	33	125	0.42	100
40	His	IDE	CON	38	125	0.42	101
50	Geo	IDE	CON	39	125	0.42	102
74	Mat	RES	PRO	34	125	0.42	103
11	Esp	INF	CON	40	120	0.40	104
81	Mat	RES	CON	34	119	0.40	105
93	Fis	RES	PRO	37	119	0.40	106
12	Esp	RES	CON	28	113	0.38	107
36	Lit	IDE	FAC	34	113	0.38	108
66	Mat	RES	PRO	29	111	0.37	109
56	Geo	IDE	CON	30	106	0.35	110
7	Esp	INF	CON	33	104	0.35	111
63	Mat	IDE	CON	30	104	0.35	112
14	Esp	IDE	CON	32	102	0.34	113
103	Quí	INF	CON	24	99	0.33	114
19	Esp	INF	CON	33	96	0.32	115
41	His	IDE	FAC	25	89	0.30	116
23	Esp	INF	CON	22	79	0.26	117
17	Esp	IDE	CON	23	76	0.25	118
116	Bio	INF	CON	21	59	0.20	119
24	Esp	INF	CON	27	30	0.10	120
Media				34.03	116.35	0.39	