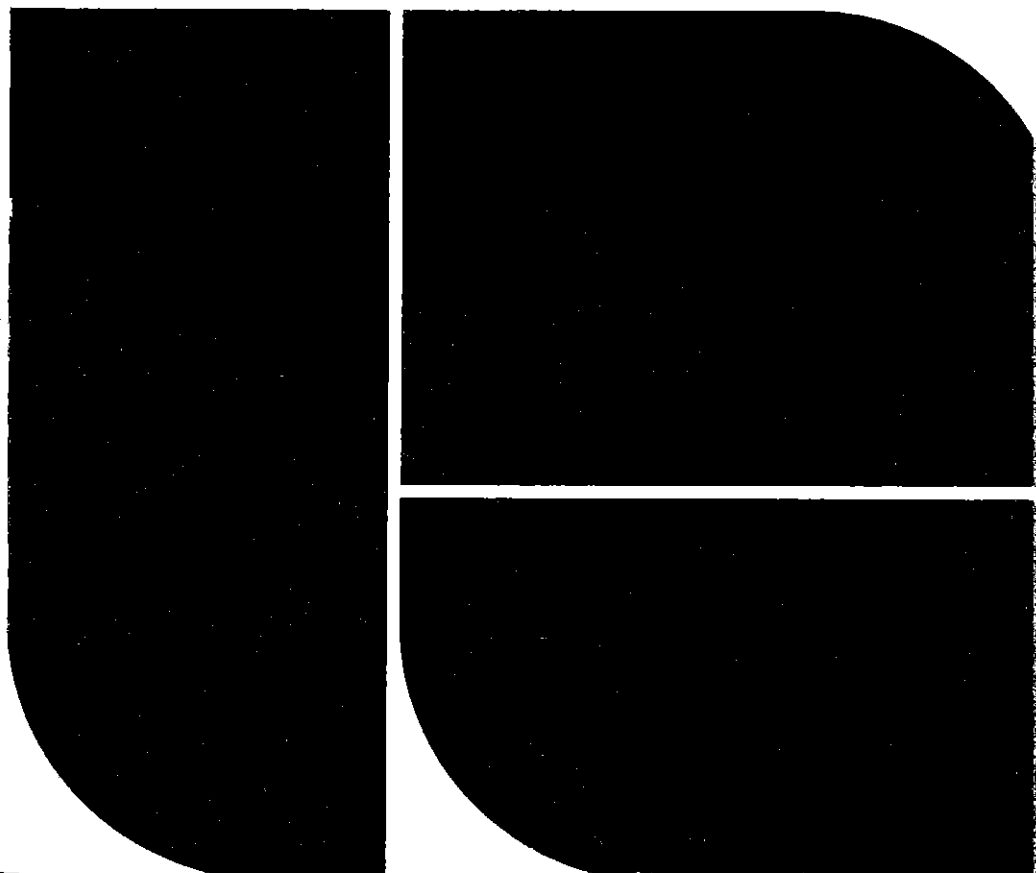


GILBERT DE LANDSHEERE

EVALUACION CONTINUA Y EXAMENES

**MANUAL DE DOCIMOLOGIA:
UN NUEVO DESARROLLO CONCEPTUAL
Y PRACTICO DE AMBOS PROCESOS**



**EVALUACION CONTINUA
Y EXAMENES**

**BIBLIOTECA NUEVAS ORIENTACIONES DE LA EDUCACION
PROBLEMATICA DE LA EDUCACION**

GILBERT DE LANDSHEERE

EVALUACION CONTINUA Y EXAMENES

Manual de docimología:
un nuevo desarrollo
conceptual y práctico
de ambos procesos



LIBRERIA "EL ATENEO" EDITORIAL

BUENOS AIRES - LIMA - RIO DE JANEIRO - CARACAS - MONTEVIDEO
MEXICO - BARCELONA - BOGOTA

Biblioteca "Nuevas Orientaciones de la Educación"
Director: Doctor Angel Diego Márquez; Asesoras
técnicas: Doctora María Arsenia Tula y Profesora N. Ruth Bähler.
Título original: Évaluation Continue et Examens. Précis de Docimologie
Traducida del francés con la autorización de las Ediciones Labor, Bruselas
© Éditions Labor-Bruxelles
Traducción de Juan Jorge Thomas.

Reservados todos los derechos. Prohibida la re-
producción total o parcial sin autorización
escrita del editor.

Queda hecho el depósito que previene la
ley. © 1973, "El Ateneo" Pedro García
S.A. Librería, Editorial e Inmobiliaria,
Florida 340, Buenos Aires.

IMPRESO EN LA ARGENTINA

INTRODUCCION

Nuevas formas de evaluación de los alumnos comienzan a penetrar paulatinamente en nuestras costumbres pedagógicas. Los resultados de la observación continua han empezado a imponerse, cada vez más, al examen de fin de año.

Tal innovación es laudable. No refleja, en modo alguno, un deseo de suprimir la medición en la escuela, sino, por el contrario, de hacerla más justa desde un punto de vista ético y más exacta científicamente.

Ese cambio sobrepasa en gran medida los límites de la calificación y coincide con transformaciones profundas de la educación. No se trata de un acontecimiento fortuito. La civilización contemporánea y la economía de nuestra sociedad requieren un hombre provisto de cualidades y conocimientos nuevos; en tal situación, las innovaciones pedagógicas se imponen en forma imperiosa.

La oposición a los exámenes tradicionales ha creado, sin embargo, un malentendido grave. Este se parece, de una manera sorprendente, al que se registró sobre todo entre 1920 y 1940, cuando la adopción de las ideas, mal dirigidas, de la pedagogía funcional de los Claparède, Dewey y otros "progresistas" condujo al culto de la improvisación, al desprecio de la disciplina rigurosa, al enfoque de intereses que solo existían en la mente de los teóricos. . .

Con el advenimiento de J. Dewey, la escuela hubiera podido elegir como lema: "¡El esfuerzo ha muerto, viva el esfuerzo!" De la misma manera nos sentimos tentados de exclamar hoy: "¡Los exámenes mueren, los exámenes están muertos, vivan los exámenes!"

Sin duda, deben desaparecer los viejos exámenes que envenenaban la atmósfera y la materia de ciclos enteros de estudios y decidían en pocas horas — ¡e incluso en pocos minutos! — sobre una carrera escolar o profesional. Indudablemente, asimismo, la mayor parte de los trabajos docimológicos publicados hasta la fecha no han sido sino ajustes de sistemas perfeccionados para seguir haciendo mal las cosas, pues no ponían en tela de juicio la enseñanza en su totalidad.

En cambio, si en cada actividad de enseñanza básica se aprovecha en debida forma el lugar que corresponde a la evaluación, al feedback* como dicen los psicólogos actuales, entonces los exámenes existen de una manera poco menos que permanente en la vida escolar. O bien consisten en evaluaciones breves o adoptan la forma de pruebas más largas y, a veces, abarcan cantidades considerables de materias.

Lo esencial es que no penetran a la manera de cuerpos extraños en el proceso educativo, sino que forman parte de él. Además, por esa misma razón, no se confinan estrechamente al dominio del saber, sino que aspiran a captar la personalidad total.

Toda reforma de fondo va acompañada, casi necesariamente, de un período de adaptación, de tanteo, de yerros. De este modo se explica la confusión de tantos maestros que realmente tratan de poner en práctica el sistema de evaluación continua.

Así se explica también que, en forma repentina, recaiga un tabú sobre la palabra examen y que, en el momento en que escribimos estas líneas, se entreguen en algunos lugares certificados de fin de estudios sobre la base de evaluaciones por puntos, cualitativa y cuantitativamente insuficientes, por añadidura mal integradas, y basadas en impresiones personales en comparación con las cuales los viejos exámenes napoleónicos parecen monumentos a la objetividad. Pero estamos convencidos de que solo se trata de una crisis de adaptación.

* Feedback: realimentación. En inglés en el original (N. del T.)

Por lo tanto no es paradójico, ni mucho menos, que después de haber rehecho —brevemente— el proceso de los exámenes y de habernos declarado fervorosos partidarios de la observación continua y del balance de fin de ciclo escolar, consagremos la presente obra nada menos que a la organización de los exámenes. Pero el lector sabe ahora que, si bien la palabra sigue siendo la misma, el sentido ha cambiado profundamente.

* * *

Muchos educadores desconfían aún de los tests y demás instrumentos que pretenden cuantificar al ser humano. No sin razón, por otra parte, porque esas herramientas sufrieron en los últimos decenios sus enfermedades de la infancia. Y como sus usuarios no siempre han sabido compensar la imperfección de las técnicas con una gran moderación en el juicio y una profunda comprensión de las situaciones de conjunto, se cometieron muchos errores.

Pero si bien la prudencia sigue siendo necesaria, si el espíritu crítico y el sentido clínico no perderán nunca sus derechos, se han logrado, no obstante, considerables progresos: estamos en condiciones de medir de una manera satisfactoria numerosos comportamientos humanos.

En la actualidad, algunos podrían rechazar la medición objetiva, no porque carezca de validez sino porque no comprenden su economía. Negar todo valor a lo que se ignora es un mecanismo de defensa bien conocido. El aspecto matemático de los métodos por adoptar y de la literatura experimental, a menudo indigesta, complican aún más la situación.

La época del antagonismo entre los partidarios de la psicometría y los educadores ha tocado a su fin. Los primeros deben humanizar sus números; los segundos, introducir más rigor en sus procedimientos; los dos han de unir y armonizar sus esfuerzos en beneficio del estudiante y de la comunidad.

* * *

Que se tranquilicen quienes, al hojear este libro, hayan visto algunos números: ¡las cuatro operaciones aritméticas les resultarán suficientes!

Como sucede siempre, los procedimientos de que se trata son más difíciles de explicar que de aplicar.

Es cierto que no se revelarán todos los secretos de la medición y la estadística, pero las nociones sencillas que se encontrarán han demostrado su valor en la práctica. Poseen, además, un innegable poder de demistificación y facilitarán, más adelante, la lectura de trabajos docimológicos más especializados.

INDICE GENERAL

INTRODUCCION	VII
PRIMERA PARTE — Definiciones	1
SEGUNDA PARTE — Acusación y defensa	9
Capítulo 1: Crítica de los exámenes	11
Capítulo 2: Defensa de la nota subjetiva y del examen	38
TERCERA PARTE — La construcción del examen	47
Capítulo 1: El objeto y los objetivos	49
I. El objeto	49
II. Los objetivos	53
Capítulo 2: Redacción de las preguntas	73
I. Observaciones generales	73
II. ¿Respuestas abiertas o cerradas?	80
Capítulo 3: La calificación	95
I. Un preámbulo indispensable: la curva de Gauss	95
II. La calificación subjetiva: escala de evaluación	107
III. La calificación objetiva	135
IV. Estandarización o medida de la posición relativa	135
Capítulo 4: Control de la fidelidad del examen	144
Capítulo 5: Control de la validez	147

XII

INDICE GENERAL

CUARTA PARTE	— Los procedimientos de moderación	153
Capítulo 1:	Planteo del problema	155
Capítulo 2:	Algunos sistemas de moderación de exámenes	161
I.	La fórmula más liberal: el sistema sueco	161
II.	Sistema de moderación establecido a partir de un <i>test</i> de conocimientos	164
III.	Moderación espontánea mediante apelación a un "banco" de preguntas	165
IV.	Un sistema sencillo de moderación global al servicio de la selección	166
V.	En Inglaterra, un sistema completo de moderación	169
VI.	La nota de fin de año	181
QUINTA PARTE	— <i>El mito de la curva de Gauss</i>	187
Capítulo 1:	Evolución de la curva de conocimientos	190
Capítulo 2:	Una pedagogía de la curva en J	194
Capítulo 3:	La teoría de la evaluación formativa	197
CONCLUSIONES Y RECOMENDACIONES		203
APENDICES		
I.	Estudio comparado de una pregunta de examen	211
II.	Redacción en lengua materna	217
III.	Ejemplo de enseñanza semiindividualizada	226
BIBLIOGRAFIA		241
INDICE ANALITICO		245

3
5
1
1
4
5
6
9
1
7
0
4
7
3
1
7
6
1
5

PRIMERA PARTE

DEFINICIONES

I. DOCIMOLOGIA, DOCIMASTICA Y DOXOLOGIA

La *docimología* es una ciencia cuyo objeto es el estudio sistemático de los exámenes, sobre todo de los sistemas de calificación y del comportamiento de los examinadores y examinandos.

La *docimástica* es la técnica de los exámenes.

En un comienzo, la docimología mostraba una actitud negativa al criticar los modos de calificación y poner de relieve, en forma experimental, la falta de fidelidad y validez de los exámenes.

Más adelante entró en una etapa constructiva tratando de proponer métodos y técnicas de medición más objetivos o, por lo menos, más rigurosos e ideando medios de hacer comparables las notas con el propósito de lograr una mayor justicia en la escuela.

La *doxología* es el estudio sistemático del papel que la evaluación desempeña en la educación escolar.

J. Guillaumin¹ le asigna, sobre todo, los siguientes objetivos:

- El estudio de los efectos inhibidores o estimulantes de las distintas formas de exámenes.
- El estudio de las reacciones emocionales de los alumnos y, por consiguiente, las reacciones intelectuales a los juicios del maestro.

¹ J. Guillaumin. "L'aspect interpersonnel de la notation: de la docimologie à la doxologie pédagogique", en *Bulletin de la Société A. Binet et T. Simon*, 86, 1968, págs. 250-275.

- El estudio de los efectos de la opinión del maestro, con respecto a sus alumnos, sobre su enseñanza y el aprendizaje escolar.
- El estudio de los procesos puestos en juego y de los efectos obtenidos por la automatización, la intercalificación, la calificación de equipo y la falta de calificación.

II. EXAMENES Y CONCURSOS. OBSERVACION Y EVALUACION CONTINUAS

El examen y el concurso difícilmente pueden separarse de la idea de *prueba*, que, por lo demás, se convirtió en su sinónimo, después de sus connotaciones más generales de sufrimiento, desgracia y peligro que traen a la luz el valor y la resistencia.

En el examen, la admisión se determina por medio de una nota que el candidato debe alcanzar o sobrepasar, mientras que en el concurso, el número de puestos ofrecidos queda fijado de antemano. En ambos casos es innegable que existen la *amenaza* y el *peligro*, el temor al rechazo. En cuanto el procedimiento está contaminado de imperfecciones graves, por pocas que sean, no estamos muy lejos de las ordalías de la Edad Media.

En la noción de concurso y examen percibimos una *carga agresiva* de la cual carecen por completo los conceptos de observación y evaluación continuas. La serenidad, la indulgencia, incluso la benevolencia, animan al maestro que sigue, con una simpatía que no excluye del todo la severidad, el largo camino de sus alumnos hacia el equilibrio en el presente y el acceso pleno al estado adulto en el futuro.

Los exámenes marcan los puntos finales de ciertas etapas; los concursos abren las puertas a los elegidos: constituyen mojones en el camino, acontecimientos en el proceso educacional.

El aprendizaje que los precede consiste fundamentalmente en una sucesión continua de comportamientos y *feedbacks*, es decir, de informaciones que permiten ver su validez, su pertinencia. No nos incumbe entrar aquí en una larga discusión acerca de las modalidades y los efectos del refuerzo

de las conductas. Lo importante es que, sin ese refuerzo, no es posible que exista, aparentemente, aprendizaje.

La evaluación, en el sentido restringido que le otorgamos en esta obra, merece, pues, un lugar importante en la enseñanza de la cual forma parte. Siempre se relaciona, en forma directa o indirecta, con el progreso, en extensión o en calidad, del aprendizaje.

La evaluación desempeña tres papeles:

1º Un papel de pronóstico: ¿posee el alumno las cualidades intelectuales y caracterológicas y los conocimientos necesarios para abordar una materia nueva o un ciclo de estudios superiores? ¿Se halla en el nivel que le corresponde? Responder a estas preguntas equivale a predecir el éxito en la etapa por iniciarse.

2º Un papel de medición:

a) Control de las adquisiciones.

b) Evaluación del progreso; en este caso se compara al alumno consigo mismo.

c) Situación del alumno en un momento dado:

- en su división o grupo de trabajo;
- en el conjunto de las divisiones paralelas de una misma escuela;
- dentro de conjuntos más amplios: ciudad, distrito, provincia, país.

No se trata necesariamente de llevar a cabo un examen o concurso, sino de precisar con exactitud el punto en que se está, de determinar la posición relativa.

3º Un papel de diagnóstico:

¿Por qué no ha sido perfecto el aprendizaje? ¿Cuáles son las materias o técnicas que el estudiante no domina del todo y cuáles son los procesos mentales en cuestión?

Los instrumentos imprescindibles para la evaluación, el examen o el concurso no siempre difieren, pero sí la forma de aplicarlos. Por esta razón, con el fin de evitar fastidiosas precisiones verbales, emplearemos en las páginas siguientes la palabra examen en dos sentidos diferentes, siempre aclarados por el contexto: el examen propiamente dicho y, de una manera más general, todo procedimiento pedagógico cuya finalidad consiste en medir un aprendizaje o un conocimiento.

Exámenes internos y exámenes externos

En sentido estricto, el *examen interno* en una materia lo construye y califica el maestro que la enseñó y lo dan los alumnos que recibieron esa enseñanza en el ámbito de la clase o de la escuela.

En un sentido más amplio, denominamos internos a aquellos exámenes organizados independientemente en cada escuela, exista o no coordinación o unificación por materia, niveles y secciones.

Exámenes externos son las pruebas organizadas y calificadas por mesas examinadoras independientes de las escuelas, a nivel local, regional o nacional. Las más conocidas de esas pruebas son las del bachillerato francés. Cabe mencionar, además, los exámenes cantonales de Bélgica, al terminar los estudios primarios, y en Inglaterra, hasta hace poco, la *Eleven + Examination* en el momento de ingresar en la enseñanza secundaria.

III. NOTAS Y "SCORES"*

La distinción entre notas y *scores* prestaría, en nuestra opinión, un gran servicio a la educación.

En un dictado, el alumno puede cometer cierto número de errores y apenas si nos incumbe evaluarlos: están o no están. Sin embargo, el número de errores encontrados no posee en sí mismo ningún significado educativo: cometer cinco errores en un dictado de un texto de Merimée señala un conocimiento extraordinario de los arcanos de la ortografía; en otras circunstancias, el mismo resultado indica una deficiencia grave. En consecuencia, necesitamos asimismo una información relativista.

Por *score* comprendemos los resultados objetivos obtenidos en un *test* o en cualquier otra forma de evaluación, mediante un *recuento o descuento de puntos según reglas fijas*: el número de faltas en un dictado, los resultados en un *test* estandarizado.

La *nota* constituye una apreciación sintética que refleja la evaluación de una ejecución en el campo educativo.

**Score*: puntaje, nota. En inglés en el original. (N. del T.)

La nota puede ser objetiva o subjetiva, pero siempre es relativa. Atribuir la nota A a un alumno cuya ejecución se halla en determinado nivel de una estandarización nacional corresponde a la primera categoría; calificar de *bien* su composición, está dentro de la segunda.

SEGUNDA PARTE

ACUSACION Y DEFENSA

CRITICA DE LOS EXAMENES

Una crítica detallada de los exámenes tradicionales fue realizada por H. Piéron en su excelente obra *Examens et docimologie*¹. Todo educador debería leerla.

El presente capítulo contiene dos clases de observaciones. Unas son breves recuerdos de imperfecciones bien conocidas, incansablemente señaladas por los docimólogos; las otras se refieren a fenómenos no tan estudiados: estereotipia, efecto de halo, efecto edípico de la predicción. . . A veces se pone en tela de juicio el examen únicamente; otras veces, toda la pedagogía a la cual pertenece.

1. Cuerpos extraños en la educación al servicio de una pedagogía anticuada.

Se trate de interrogatorios periódicos, o bien de exámenes trimestrales o anuales, la evaluación se reduce muchas veces a controlar la retención de conocimientos y deja sin explorar no solo los aspectos más importantes de la inteligencia, sino también casi todos los rasgos de la personalidad que una educación bien comprendida debe cultivar.

De hecho, estamos frente a las consecuencias de un sistema pedagógico en el cual, como suele afirmar T. Brameld, las clases funcionan a la manera de cintas transportadoras de conocimientos y de valores, seleccionados en función de un papel predeterminado por desempeñar en una sociedad no democrática.

¹ París, P.U.F., 1963.

Dentro de esos márgenes, el examen constituye una especie de control de fabricación, de verificación de la conformidad al molde, al patrón; en suma, a las especificaciones impuestas por la autoridad.

La evolución de la persona humana apenas si entra en esas preocupaciones, porque la educación no se concibe con ese propósito, sino que se halla al servicio de un régimen.

Las infaustas consecuencias de tal situación se analizaron ya en muchas oportunidades. En vez de servir al alumno, de informarlo de un modo funcional acerca del valor de su desempeño en el trascurso del aprendizaje, de crear una mejor adaptación y, por consiguiente, de ofrecer una evaluación aceptada a tal punto que el estudiante participe en ella con sinceridad y en forma espontánea, el examen es tan mal recibido como una declaración de impuestos, que, ciertamente, tampoco es generadora de civismo.

Además, el profundo rechazo del procedimiento, en combinación con la pobreza intelectual de las preguntas planteadas, lleva a un resultado fácil de imaginar: no es excepcional que en los quince meses que siguen a un examen se olvide un 80% de las materias fácticas.

De esta manera, si dejamos de lado la función social del examen, el fracaso es total: no hemos educado, ni instruido.

2. Ansiedad y stress*

En una civilización en la que el éxito escolar es condición previa del triunfo material y social, a tal punto que Young se creyó en el deber de denunciar los peligros de la "meritocracia", el examen que decide sobre el paso al año siguiente o la obtención del diploma es temido por el niño y su familia. Incluso en el trascurso del año, las pruebas de control de conocimientos y aptitudes se abordan con tensión y desconfianza; esto, por cierto, no es un estado ideal y desnaturaliza profundamente el papel educativo de la medición del aprendizaje.

Agrava aún más el cuadro el hecho de que nuestra tradición impone que los exámenes "grandes", que abarcan

* *Stress*: término general que designa una agresión, una acción violenta ejercida sobre un organismo. (N. del T.)

todas las materias, se lleven a cabo sucesivamente en pocos días. En oportunidad de un período lectivo en la universidad tuve, un año, el triste privilegio de ser interrogado en catorce materias en dos días consecutivos. . . En la enseñanza secundaria, el mismo fenómeno de concentración se produce muchas veces durante el año cuando, al acercarse la fecha de presentar el boletín periódico a los padres, se plantean una cantidad de interrogantes.

En ambos casos, la acumulación de las pruebas y su preparación, obstaculizadas a menudo por la introducción de materias nuevas hasta el último momento (los profesores no siempre respetan los períodos de revisión o no los aprovechan de la manera más adecuada), fatigan considerablemente el cuerpo y la mente.

Algunos pretenden que esas condiciones difíciles poseen en sí mismas un valor educativo: la vida moderna no nos evita el *stress* y conviene estar preparado para él. De acuerdo, siempre que las cosas se presenten con toda claridad.

Antes de organizar un examen es importante que se defina abiertamente su objetivo: ¿se trata de controlar la adquisición de conocimientos, la resistencia al *stress*, o bien la capacidad de restituir y utilizar los conocimientos en situación de *stress*? Son tres objetivos diferentes que requieren pruebas diferentes.

3. Desigualdad - injusticia

En nuestro sistema escolar, los profesores redactan personalmente las preguntas de examen para sus alumnos. El principio es excelente, y en países como Gran Bretaña, donde por tradición todos los exámenes clave estaban a cargo de comisiones externas a la escuela, se tiende a delegar en cada docente la responsabilidad del control. Sin embargo, los métodos anglosajones son más rigurosos que los nuestros porque se toman medidas precisas a fin de asegurar la comparanza de los exámenes en la región, e incluso, en todo el país. Dedicamos parte de este libro a los procedimientos empleados para lograr ese objetivo.

La libertad, poco menos que total, que tienen nuestros maestros conduce a situaciones de injusticia.

Las escuelas de nuestro país constituyen mundos aislados cuyas poblaciones presentan características a veces muy diversas. Un pequeño establecimiento de enseñanza secundaria cuenta con unos pocos alumnos, en su mayoría disminuidos por su origen socioeconómico. Otro establecimiento, en cambio, radicado en una gran ciudad, ve sus divisiones "fuertes" pobladas por una mayoría favorecida social e intelectualmente, porque los adolescentes con dificultades son orientados hacia una escuela vecina que goza de la reputación de ser más indulgente.

Con el tiempo, y con ayuda de la competencia, comienzan a crearse tradiciones de severidad o de "generosidad" en los establecimientos. Marion Coulon observó, en fecha reciente, que cuando se crean sucesivamente escuelas de un mismo tipo en una misma región, la severidad de los exámenes se halla en proporción inversa a la antigüedad del establecimiento. No podemos generalizar, por cierto, pero, ¿quién se atrevería a afirmar que tal proposición carece de todo fundamento? Así lo comprendieron muchos padres que ambicionan un determinado diploma para su hijo, sin preocuparse demasiado por el valor intrínseco de los estudios; si fracasa, es más fácil cambiar de escuela que recurrir a un tratamiento o modificar la orientación.

Sea como fuere, los maestros adaptan su enseñanza y los exámenes al nivel de su clase (y esto es digno de elogio), con la consecuencia de que un alumno bien calificado en una población débil se encontraría entre los muy débiles o fracasaría en un grupo fuerte.

El efecto acumulativo de todas las adaptaciones circunstanciales desemboca a veces en situaciones escandalosas. Hace algunos años consultamos unas estadísticas según las cuales, para un mismo tipo y nivel de enseñanza secundaria, las perspectivas de éxito de un alumno variaban entre uno y ocho puntos según la escuela frecuentada. Constituye un caso extremo, probablemente, pero una vez más nos faltan estudios exactos que nos permitan saber lo que realmente sucede.

Cabe insistir en el hecho de que la adaptación del maestro al nivel de sus alumnos no es criticable en sí misma; por el contrario, es una de las características del buen educador.

La injusticia aparece cuando notas puramente relativas se utilizan en una situación de competencia externa, o bien

cuando constituyen el único criterio para la obtención de un diploma del cual puede depender el porvenir de los alumnos.

De todos modos, comprobamos variantes considerables en cuanto al *número de materias incluidas* y la *calidad de las respuestas exigidas*. En algunos lugares basta con la repetición de memoria de una parte de un breve resumen; en otras partes, las preguntas exigen análisis, síntesis y juicio personal respecto de un abundante material.

Un maestro hace solamente una o dos preguntas, elegidas más bien por la facilidad de corrección que por su verdadera importancia, mientras que otro se esfuerza por recorrer todo el programa, con el peligro de convertir el examen en una maratón. En el deporte es sabido que esa larga carrera ha de reservarse a individuos excepcionales; ¿por qué, en el campo de la instrucción, los treinta alumnos de una clase deberían ser todos corredores de fondo?

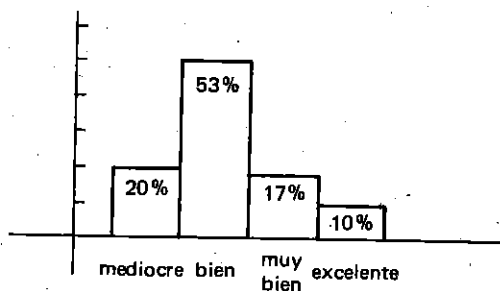
Antes de proseguir, tres resultados de investigación nos permitirán ejemplificar los fenómenos que venimos tratando.

R. Gjorgjevski² mostró, mediante un simple experimento, cómo los profesores evalúan de diferentes maneras, desde un punto de vista relativo, un mismo trabajo.

Cinco correctores, profesores de la misma materia, calificaron independientemente uno de otro cien pruebas escritas provenientes de un mismo colegio secundario. Luego se extrajeron quince trabajos que habían obtenido la nota "bien". Estos se entregaron para una nueva corrección a otros cuatro profesores, quienes adoptaron, en forma espontánea, nuevas exigencias: solo en la mitad de los casos, aproximadamente, la nota seguía siendo "bien". Veamos la nueva distribución³ en la página siguiente.

² Véase N. Rot y Z. Butas, "Les distributions des notes scolaires comparées aux distributions des résultats obtenus aux tests de connaissances", en *Le travail humain*, XXII, 1-2, 1959.

³ El número de trabajos es demasiado reducido como para conducir a resultados estadísticamente significativos.



Siguiendo con su experimento, Gjorgjevski extrajo, de la misma serie de cien trabajos, un grupo de doce considerados muy buenos y otra docena juzgados mediocres.

En cada grupo introdujo tres trabajos calificados como buenos, evaluados dos veces por cinco profesores. La media de los trabajos "buenos" mezclados entre los "muy buenos", descendió, de 3 sobre 5, a 2,40; en el grupo opuesto, la media pasó de 3 a 3,87.

Por otra parte, debemos a Reuchlin⁴ un estudio referente a 4.860 alumnos de 397 escuelas primarias de Francia. Se invitó a los maestros a calificar a sus alumnos en cuatro categorías: muy buenos, buenos, regulares y mediocres. Después se sometió a una misma prueba de francés y aritmética a todos los alumnos. El examen de los resultados mostró que se hallaban en un mismo nivel de conocimientos los alumnos juzgados por la escuela como muy buenos, buenos, regulares o mediocres.

Tal como señala acertadamente Reuchlin: "Puede afirmarse (...) que el maestro conoce ciertamente mejor que nadie los puntos del programa que cada uno de sus alumnos tiene asimilados o no. Lo que ignora es el peso que se atribuye a cada deficiencia, a cada laguna, cuando se la considera, no ya en el conjunto de una clase, que puede ser "fuerte" o "débil", sino con relación al país en su totalidad. A ello se deben las divergencias de apreciación puestas de relieve por la encuesta".

⁴ Véase *Le travail humain*, XXII, 1-2, 1959, pág. 12 y sig.

Los siguientes gráficos sintetizan los resultados

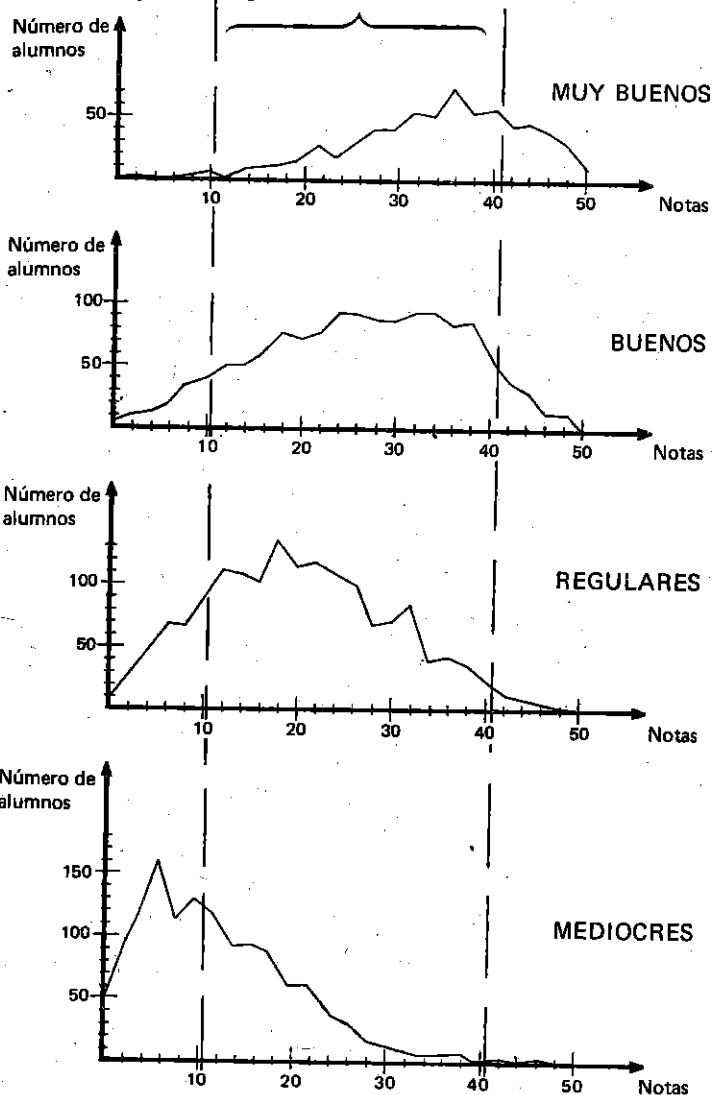


Fig. 1. — Los cuatro gráficos permiten comparar la distribución de las notas obtenidas en la prueba de cálculo por 654 alumnos juzgados como "muy buenos" por sus maestros, por 1.303 alumnos "buenos", 1.551 "regulares" y 1.300 "mediocres". Vemos que las distribuciones de las notas se superponen ampliamente: en la zona de notas que va de 10 a 40, el mismo nivel de conocimientos puede calificarse, según el caso, de muy bueno, regular o mediocre.

He aquí dos ejemplos más proporcionados por F. Bacher⁵ :

“En oportunidad de una encuesta realizada en Francia, y referida a una muestra representativa de alumnos de tercer año (fin del primer ciclo secundario), se tomó una prueba de conocimientos literarios y otra de matemáticas a los alumnos de 406 divisiones. Se comprobó que los promedios por división se escalonaban de 23 a 60 en francés (una prueba calificada sobre 80 puntos), y entre 7 y 38 en matemáticas (una prueba calificada sobre 44 puntos) (Reuchlin, Bacher, 1968). (...) En Estados Unidos, Flanagan (1964) señala que alumnos clasificados en el cuarto inferior de su escuela estarían en el cuarto superior si frecuentaran otras escuelas de la misma zona.”

4. El fracaso, generador de fracasos

Podría objetarse que los estudios superiores o la vida misma se encargarán de restablecer la justicia que las notas escolares hubieran violado. Lamentablemente, las cosas no son tan sencillas. En primer término, porque los alumnos eliminados o injustamente desanimados frente a la enseñanza secundaria en general, por ejemplo, tal vez ya no tengan oportunidad de defenderse en la universidad. Por otra parte, los favoritos pueden “ganar en todos los tableros”.

Las investigaciones modernas confirman un viejo refrán pedagógico: el éxito engendra el éxito y el fracaso actual prepara otro en lo futuro. Llevados por el éxito, algunos alumnos valorizan un capital intelectual limitado. El paso “barato” de un grado a otro trae consigo, a veces, la doble consecuencia saludable de no incitar a los padres a retirar al hijo de su escuela y de permitir al estudiante encontrar una especie de nuevo aliento. ¿Cuántas veces no ven los profesores secundarios que el rendimiento escolar de un alumno se levanta en forma espectacular una vez que la crisis física y psicológica de la adolescencia pierde parte de su agudeza?

⁵ F. Bacher, “La normalisation de la notation”, en *Docimologie et Education*, número especial de la revista *Les Sciences de l'Education*, 2-3, 1969, págs. 51-52.

R. Rosenthal y L. Jacobson⁶ han publicado en fecha reciente un estudio de conjunto sobre el aspecto del *efecto edípico de la predicción*, denominado de esa manera porque si el oráculo, al nacer Edipo, no hubiera predicho que mataría a su padre, no hubiera sido alejado de su familia. De conocer a su padre, no lo hubiera matado. En otras palabras, la tragedia sobrevino *por haberse profetizado*.

Aparentemente, en una medida difícil de apreciar, por cierto, un alumno se comporta en función del juicio que el profesor se formula acerca de él. Ni siquiera los animales domesticados escapan a esa regla. Rosenthal recuerda el siguiente experimento: se forman dos grupos de ratones blancos genéticamente idénticos. En el momento en que se entregan a los estudiantes encargados de enseñarlos, se les hace la observación de que el primer grupo se compone de animales particularmente bien dotados, en tanto que el segundo es de poca calidad. Los resultados de la enseñanza confirmarán ese pronóstico . . . fantaseado.

El siguiente experimento, realizado también por Rosenthal, pertenece a la misma categoría.

Anunció a los maestros de los seis grados primarios de la Oak School que unos eminentes investigadores habían preparado un *test* de "desenvolvimiento" (en realidad se utilizó un simple *test* de inteligencia⁷, poco conocido y sin ninguna virtud especial).

Se aplicó a todos los alumnos de la escuela y se señalaron aquellos casos que estaban a punto de desenvolverse intelectualmente, respondiendo a un pronóstico de pura fantasía.

El cociente intelectual de los alumnos elegidos de esa manera se elevó de un modo significativo en los tres años inferiores; además, se observaron progresos superiores al promedio en lectura y aritmética⁸.

⁶ *Pygmalion in the Classroom, Teacher Expectations and Pupil's Intellectual Development*, New York, Holt, Rinehart and Winston, 1968.

⁷ Medido mediante el Toga de Flanagan (1960).

⁸ Los números presentados por Rosenthal y Jacobson fueron impugnados por R. L. Thorndike. En cambio, la tendencia al incremento de los CI y los puntajes ha sido reconocida aparentemente por todos los especialistas. Sin duda, esto es lo que importa. Para más detalles, véase *American Educational Research Journal*, 5, 4, 1968, 708.

5. Ruptura entre enseñanza y examen

El examen ha de reflejar la enseñanza impartida. Imaginémonos que un profesor de educación física, en vez de llevar a los alumnos a la piletta, hubiera dedicado el año a disertar sobre natación. Sería evidentemente inadmisibile que las notas del examen se dieran según las actuaciones natatorias... , porque solo aprobarían aquellos que hubieran aprendido a nadar fuera de la escuela.

Semejante proceder parece inimaginable. Sin embargo, cuántas veces observamos una aberración similar. El profesor dicta un curso durante el cual impone informaciones, opiniones y juicios; después, en el examen, hace preguntas denominadas "de comprensión", es decir, aquellas que harían poner en práctica capacidades que precisamente no fueron engendradas durante las clases. ¿Por qué los estudiantes han de ser capaces de descubrir de repente, por sí solos, en el momento del examen, la solución de aplicaciones geométricas originales, si no pudieron hacerlo durante el año?

De esta manera se explican muchos porcentajes anormales de fracasos. Una investigación hecha a causa de la gran cantidad de "insuficientes" de un profesor de química (solo 4 alumnos de 22 habían superado la mitad de los puntos en el examen de fin de año), reveló rápidamente que el profesor no tomaba más de dos o tres pruebas escritas por año. Dos de las tres preguntas del examen requerían interpretación y transferencia, pasos que, en el nivel de dificultad en que se colocaba el maestro, necesitaban una comprensión profunda de los fenómenos y una ejercitación larga y bien controlada en la solución de los problemas originales.

Una conversación amistosa con el joven profesor en cuestión reveló que se creía obligado a tratar en detalle todos los puntos del programa, que no tenía conciencia de la necesidad de una ejercitación sistemática, y por lo menos algo individualizada, para la solución de problemas originales, y que, de todos modos, no poseía las nociones psicopedagógicas de base indispensables para llevar a cabo una ejercitación semejante.

6. Desacuerdo entre las personas encargadas de corregir la tarea escolar

Todos los docimólogos citan ejemplos de grandes divergencias entre maestros llamados a evaluar un mismo trabajo de un alumno, o un mismo conjunto de trabajos; esto permite, con todo, establecer una relatividad más segura. Las fluctuaciones de las notas de un profesor que juzga el mismo trabajo con cierto intervalo (fidelidad) pueden ser, asimismo, considerables.

Como todas las obras sobre docimología abundan en ejemplos, proponemos solo algunos para recordar.

a) *La redacción francesa.*

Recibe preferentemente las acusaciones.

- Piéron menciona que una misma redacción, juzgada por 76 educadores, todos profesores de lenguaje, llevó a los siguientes resultados⁹:

<i>Nota sobre 20</i>	<i>Número de personas encargadas de la corrección</i>
de 0 a 1	1
2 - 3	6
4 - 5	20
6 - 7	34
8 - 9	10
10 - 11	3
12 - 13	2

- El Servicio de Investigación Pedagógica del C.R.D.P. de Lyon invitó, en 1967¹⁰, a 150 profesores de francés de la Academia de esa ciudad a corregir tres redacciones francesas sobre un mismo tema.

⁹ Piéron, *op. cit.*, pág. 123.

¹⁰ C. R. D. P., *Docimologie et Examens*, Lyon, I.P.N., 1969.

<i>Máx. 20</i>	<i>Mediana</i>	<i>Margen de variación</i>
Deber I	10½	4 ½ - 13 ½
Deber II	6,5	2 ½ - 12 ½
Deber III	11,6	5 ½ - 17 ½

¿Mejoraría la situación con la intervención de hombres cultos, extraños a los quehaceres cotidianos de la enseñanza? P. Diederich y sus colaboradores¹¹ compararon la manera de corregir de 10 profesores de lenguaje y de 43 personas no docentes, especialistas en ciencias sociales y naturales, escritores, editores, abogados y hombres de negocios (directivos).

Todos fueron invitados, sin otra indicación, a calificar 300 composiciones en 9 grupos, por orden de mérito general. Ninguno de los grupos podía contener menos de 12 trabajos (4%). Al 94% del conjunto se le atribuyeron 7, 8, ó 9 de las 9 categorías posibles. Ninguna redacción recibió menos de cinco calificaciones diferentes. . .

b) *Matemáticas*

El C.R.D.P. de Lyon¹² solicitó a 150 profesores de matemáticas de 3^{er} año que calificaran los trabajos de 3 alumnos que debían resolver el mismo problema.

<i>Máx. 20</i>	<i>Mediana</i>	<i>Margen de variación</i>
Alumno I	5,70	0,5 - 11,5
Alumno II	16	11,5 - 20
Alumno III	8	3,5 - 11,5

¹¹ P. Diederich y otros, *Factors in Judgments of Writing Ability*, Princeton, E.T.S., Res. Bulletin, RB-61-15, 1961.

¹² *Op. cit.*

Por ejemplo, si para el alumno I se eliminarían las dos notas extremas, son tantas las personas para calificar 2/20 como 8/20, cuando cabía esperar una calificación objetiva.

c) *Medicina*

A nivel universitario, los resultados no son más seguros.

Un mismo candidato de 2º año de medicina rindió, en junio y setiembre de 1965, una prueba escrita de cinco preguntas simples y precisas para contestar en 1 h 30 m¹³.

Las respuestas (anónimas) fueron corregidas independientemente por cinco personas que poseían títulos y aptitudes idénticos.

	Resultado total sobre 100	Margen de variación	Decisión ¹⁴	
			Admitidos	Rechazados
Trabajo de junio	47	- 78	3	2
Trabajo de setiembre	32,5	- 73	3	2

Si se consideran las notas atribuidas a cada una de las 10 preguntas, comprobamos:

1) Desvío máximo:	12	(de 3 a 15 sobre 20)
2) Desvío para 4 preguntas:	11	
3) Desvío máximo menor entre cinco personas encargadas de la corrección:	7,5	

d) *Varios*

A. Agazzi¹⁵ menciona un ejemplo más sorprendente aún con referencia a un conjunto de materias.

¹³ Véase *Le Monde*, 6 set. 1966.

¹⁴ Cabe señalar que los tres admitidos fueron los mismos en ambos casos.

¹⁵ A. Agazzi, *Les aspects pédagogiques des examens*, Estrasburgo, Consejo de Europa, 1967, pág. 119. El estudio pertenece a H. Laugier y D. Weinberg y se refiere al bachillerato francés.

Seis personas encargadas de la corrección calificaron, en cada caso, los exámenes del bachillerato (nota de aplazo: menos de 10 puntos sobre un total de 20):

	<i>Rechazados por los seis encargados de la corrección</i>	<i>Admitidos por los seis encargados de la corrección</i>	<i>Admitidos por los unos y rechazados por los otros</i>
Traducción al latín	40%	10%	50%
Redacción francesa	21%	9%	70%
Inglés	37%	16%	47%
Matemáticas	44%	20%	36%
Filosofía	9%	10%	81%
Física	37%	13%	50%

e) *¿Cuánta gente encargada de la corrección haría falta para estabilizar la nota?*

¿Se obtendría una mejoría considerable si se calculara el promedio entre dos de las personas que corrigen? No, si no se ha aplicado ninguna medida estricta para poner de acuerdo a los encargados de esa tarea. ¿Cuántos profesores diferentes tendrían que evaluar el mismo trabajo para lograr la estabilidad de la nota? Las siguientes cifras, indicadas por Piéron, no poseen valor absoluto, por cierto. Sin embargo, confieren una idea acerca de la magnitud de los números. . .

Redacción francesa	78
Traducción latina	19
Inglés	28
Matemáticas	13
Disertación filosófica	127
Física	16

Historia y geografía no figuran en la lista; no obstante, las dos materias dan lugar, asimismo, a grandes fluctuaciones.

Podrían multiplicarse los ejemplos de desacuerdo entre la gente que debe calificar. Pero, ¿para qué sirve complacerse con la acusación fácil, como demasiados docimólogos han

hecho? Lo importante es tener presente el peligro y, sobre todo, tratar de remediarlo.

¿Cómo se explican las divergencias observadas?

La explicación reside principalmente en la multiplicidad de los puntos de vista, de las dimensiones, según las cuales una misma prueba puede juzgarse, y en la falta de directivas precisas (a veces no existe ninguna) para la calificación. A este respecto escribe F. Bacher¹⁶:

“En una disertación es factible calificar la organización de las ideas, su originalidad, la corrección de la expresión, la elegancia del estilo, etcétera. Si se considera cada aspecto posible como una dimensión de la calificación, el trabajo de un alumno puede situarse en un espacio de n dimensiones; la dificultad surge porque, en ese espacio, no se establece sino un orden parcial entre los alumnos: un educando que ocupa una posición elevada en todas las dimensiones puede ser declarado superior a otro que ocupa un lugar menos elevado en todas ellas, pero con respecto a un alumno que ocupa una posición elevada en una primera dimensión y otra débil en una segunda, no se puede afirmar si es superior o inferior a otro que ocupa posiciones inversas.”

Además, aun considerando un mismo aspecto, quienes califican varían en cuanto a severidad, estabilidad de juicio, resistencia al efecto de halo, etcétera.

7. Inconsecuencia de una misma persona encargada de la corrección

¿Aquel que corrige será, por lo menos, fiel a sí mismo? No. A fenómenos tan evidentes como las fluctuaciones de la salud física y mental y la evolución del saber se agregan gran cantidad de factores más o menos bien definidos: variación en la calidad de la relación establecida con el alumno, en el contexto de la evaluación (si se acaba de corregir un trabajo excelente, el siguiente puede subestimarse...), en la escala adoptada de manera consciente o inconsciente, etcétera.

Catorce historiadores fueron invitados a calificar por segunda vez quince composiciones, después de un lapso de doce a diecinueve meses desde la primera calificación. Se

¹⁶ F. Bacher, *La normalisation des notes*, op. cit., págs. 53-54.

había borrado toda marca de corrección anterior. Los profesores no solo otorgaban puntos; indicaban, asimismo, el éxito general o el fracaso.

En 92 casos, de un total de 210, el veredicto difería de la vez anterior¹⁷.

Sin embargo, cabe insistir en el hecho de que resultados tan pobres se deben a la falta de directivas rigurosas que precisen los aspectos que deben considerar las personas encargadas de la corrección.

Un esquema para continuar la investigación

Resulta conveniente que, de cuando en cuando, se lleven a cabo investigaciones acerca de la fidelidad de la persona que corrige, para consigo misma o respecto de la coincidencia entre varias, aunque solo sea con el propósito de volver a llamar la atención sobre un peligro amenazador. Por esa razón creemos que es oportuno exponer el excelente plan de investigación elaborado por dos de los mejores docimólogos contemporáneos: F. Yates y D. Pidgeon.

- Se formaron siete grupos de 50 niños que terminaban la escuela primaria. Fueron sometidos a dos pruebas de lenguaje:
 - a) redacción sobre un tema elegido entre tres o cuatro;
 - b) preguntas referentes a la comprensión de textos y a la gramática (uso del idioma).
- Siete examinadores experimentados —cinco hombres y dos mujeres— calificaron los trabajos.
- Cada grupo de niños rindió, en primer término, tres exámenes diferentes con intervalos de una semana; después, la primera de las tres pruebas volvió a repetirse dos veces con un intervalo de una semana.

¹⁷ Hartog y Rhodes, *An Examination of Examinations*, Londres, McMillan, 1935, págs. 81 y 15.

ESQUEMA¹⁸

GRUPO	Turnos de examen				
	I	II	III	IV	V
1	Aae	Bac	Dab	Aac	Aae
2	Bbf	Cbd	Ebc	Bbf	Bbf
3	Ccg	Dce	Fcd	Ccg	Ccg
4	Dda	Edf	Gde	Dda	Dda
5	Eeb	Feg	Aef	Eeb	Eeb
6	Ffc	Gfa	Bfg	Ffc	Ffc
7	Ggd	Agb	Cga	Ggd	Ggd

Siete series de preguntas: A, B, C, D, E, F, G.

Siete examinadores: a, b, c, d, e, f, g.

No entraremos en detalles en cuanto a los resultados pues solo pretendíamos mostrar un ejemplo de control bien llevado, y sobre todo, con una medición del aprendizaje.

8. Estereotipos y efectos de halo

En la situación escolar habitual, el maestro conoce a cada uno de sus alumnos y, en consecuencia, está en condiciones de dosificar, matizar, sus notas en función de un efecto deseado: en un caso alienta mediante la estimación exagerada de un trabajo; en otro, muestra una severidad excepcional para causar un impacto del cual se espera un efecto saludable. En esos casos, el maestro procede deliberadamente, con plena conciencia.

Algo muy diferente ocurre con los fenómenos de estereotipia y de halo.

a) *Estereotipia*

Entendemos por *estereotipia* una inmutabilidad más o menos manifiesta en cuanto al concepto que se tiene del alumno.

¹⁸ Pertenece a D. Finney, citado por Yates y Pidgeon, *Admission to Grammar Schools*, Londres, N.F.E.R., 1957, pág. 99.

La estereotipia resulta de una contaminación de los resultados. Un primer trabajo mediocre hace pensar que el segundo también lo será. Si esto sucede, la tendencia a otorgar una nota mediocre al tercer trabajo aumenta aún más, y así sucesivamente. En un profesor sobrecargado de correcciones, la deformación se produce con mayor facilidad.

Sería un error creer que la estereotipia influye únicamente en las evaluaciones netamente subjetivas (disertación, redacción de historia y, en general, respuestas que implican un juicio de valor). Alcanza a ejercicios tan "objetivos" como un dictado de ortografía. La siguiente experiencia así lo demuestra.

Un profesor de lenguaje incluye regularmente dictados en su programa. Pronto conoce a los alumnos que suelen ser los mejores y a aquellos menos satisfactorios en ese ejercicio. Si se determina la frecuencia de los errores "pasados por alto", no advertidos por la persona que corrige, se comprueba que los olvidos a favor de los buenos alumnos son significativamente más numerosos que cuando se trata de los flojos. En el primer caso, el maestro espera no hallar errores; en el segundo, los busca¹⁹.

Si interviene una falta de conciencia profesional, la situación puede llegar a ser muy grave. Se nos había señalado un caso de estereotipia manifiesta cuya víctima era un alumno con sus traducciones latinas (ciclo secundario). Para verificar lo que sucedía, encargamos sucesivamente los trabajos posteriores a otro alumno de la clase, a un alumno de igual nivel pedagógico calificado como el mejor en otra escuela, después a un licenciado en filología clásica: la nota no varió de medio punto sobre veinte. . .

Repetimos ese tipo de experimentos en otras ocasiones. Además, cuántos padres, a veces muy competentes en la materia en que incidentalmente ayudaron a su hijo realizando la tarea en su hogar, no sintieron cierta desilusión o asombro al enterarse de "su" nota.

Sin embargo, sería injusto deducir que todos los maestros acusados de estereotipia carecen del sentido de responsabili-

¹⁹ Véase al respecto M. Zillig, *Beliebte und unbeliebte Volksschülerinnen*, en *Arch. f. d. ges. Psychologie*, 12, 1934, pág. 32, citado por E. Hohn, *Der schlechte Schüler*, Munich, R. Piper, 1967, pág. 105.

dad. Son los métodos de evaluación, los que deben ponerse en tela de juicio y no los hombres.

b) *Efecto de halo*

El *efecto de halo* presenta un marcado carácter afectivo. A menudo se sobrestiman las respuestas de un alumno de buen aspecto, con mirada franca y dicción agradable. Sin embargo, no hay que generalizar. Conocemos a profesores que, por anticonformismo, favorecen cierto desaliño u originalidades que no son siempre del mejor gusto.

Ya sea por su carácter bien legible, o por razones netamente afectivas, la letra puede influir, asimismo, en la persona que corrige. Los especialistas en publicidad saben desde hace largo tiempo que la presentación del mensaje surte un efecto considerable sobre el resultado.

Los siguientes experimentos de R. Weiss²⁰ muestran con qué facilidad una combinación de estereotipia y efecto de halo puede provocarse artificialmente.

R. Weiss seleccionó dos redacciones de alumnos de 4º grado de la escuela primaria. Los trabajos fueron mecanografiados sin modificación alguna y luego sometidos para su corrección a dos grupos de 46 maestros que también pertenecían a 4º grado.

El siguiente comentario acompañó la distribución de los trabajos entre el primer grupo: "Aquí tenemos dos composiciones escritas por alumnos de 4º grado. El trabajo n° 1 es obra de un alumno regular a quien le gusta leer las historietas dibujadas de los periódicos; los padres son empleados. El trabajo n° 2 pertenece a un alumno bien dotado; el padre es redactor de un conocido diario". Para el segundo grupo de maestros, los papeles se invirtieron.

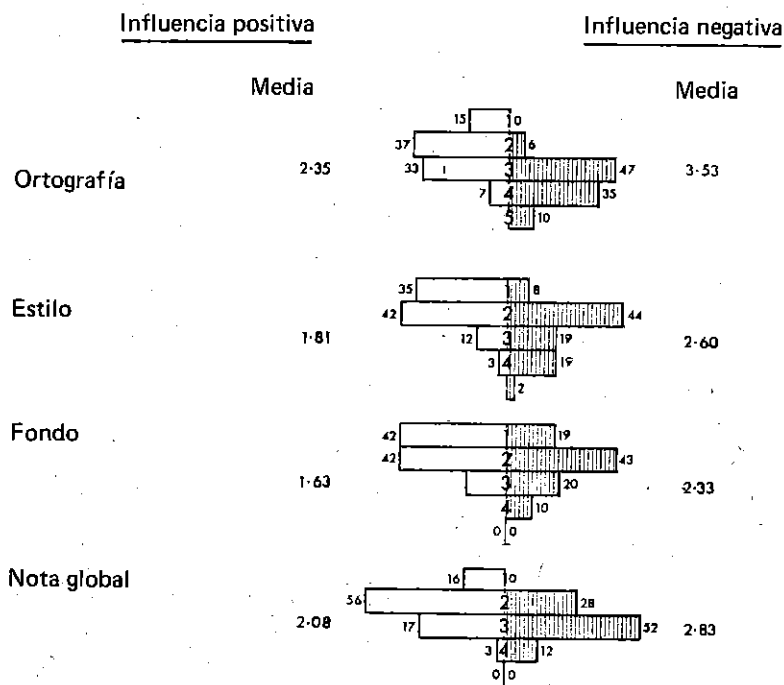
La corrección debía hacerse de acuerdo con una escala de 5 grados: muy bien (1); bien (2); regular (3); suficiente (4) e insuficiente (5). Tenían que calificarse por separado la ortografía, el estilo, el fondo y después el conjunto.

Para los cuatro aspectos considerados, las notas atribuidas al trabajo hacia el cual se había creado un prejuicio favorable

²⁰ R. Weiss, "The Reliability of the Number Marking System: An Austrian Study, en J. Lauwerys y D. Scalon, *Examinations*, Londres, Evans Brothers Ltd., 1969, págs. 101-107.

resultaron significativamente superiores. En cuanto a la ortografía, que aparentemente debía librarse del efecto de halo, se observa que al trabajo del alumno presentado como excelente, un 16% de los encargados de la corrección le otorga la nota *muy bien* y ninguno *insuficiente*. Al presentarse el mismo alumno como regular, no le fue concedido ningún *muy bien*, y el 11% lo calificó con un *insuficiente*.

CALIFICACION DE LAS REDACCIONES POR INFLUENCIA DE UN PREJUICIO FAVORABLE O DESFAVORABLE²¹



²¹ Todas las diferencias son estadísticamente significativas en P. 01, menos una (P. 10).

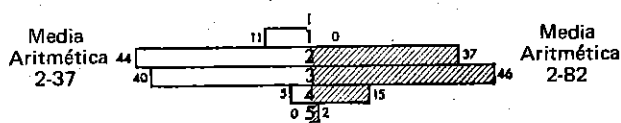
Respecto de un experimento similar referido a problemas de aritmética, también en un 4° grado, R. Weiss observa:

<i>Prejuicio favorable</i>	<i>Prejuicio desfavorable</i>
11% muy bien	ningún muy bien
5% suficiente	15% suficiente
0% insuficiente	2% insuficiente

PROBLEMA DE ARITMETICA

Influencia positiva

Influencia negativa



Es fácil advertir la influencia que ya ejerce un artificio relativamente rudimentario. Tal como subraya Weiss, la deformación será probablemente mucho mayor aún en aquella situación real en la que el maestro conoce el aspecto exterior del alumno, su conducta, su manera de expresarse, la profesión y la posición social de los padres, etcétera.

9. Falta de validez

Los exámenes tradicionales no solo son deficientes como medios de inventario; su valor de pronóstico resulta, asimismo, dudoso en muchos casos.

Junto con el inspector de uno de los mejores colegios secundarios de Bélgica observamos que dos categorías de estudiantes sobresalían brillantemente en la enseñanza universitaria: aquellos que durante toda su escolaridad parecían burlarse de las dificultades y permanecían a la cabeza de su división, y una parte no despreciable de aquellos que casi nunca habían sufrido un fracaso, pero, sin embargo, se mantenían apenas por encima de la nota fatal. En ambos casos, los rendimientos en el ciclo secundario son de buen augurio, no tanto por su índole manifiesta como por la

facilidad de adaptación, por la plasticidad intelectual que conduce a ellos.

En consecuencia, resulta bastante azaroso pronosticar el éxito o el fracaso en la facultad sobre la base del porcentaje obtenido en el momento de finalizar el ciclo secundario.

10. Un instrumento de inmovilidad social

El examen tradicional se justifica, por lo menos, tanto como medio de selección que como instrumento de promoción de la educación individual. Además, esa selección, en muchos casos, no se realiza tanto en función de la calificación técnica como de la social.

De esta manera, el examen se convierte en instrumento preferido de la inmovilidad social, fenómeno analizado en forma penetrante por J. C. Passeron, autor del que extraemos lo esencial de las siguientes consideraciones²².

a) *Efectos irreversibles del certificado escolar*

Una vez obtenido, un título escolar sigue al individuo durante toda su vida, le asegura *grosso modo* su lugar en la jerarquía profesional, su nivel de remuneración y el poder de que dispondrá.

Ahora bien, en el momento de obtener un diploma, "la competencia se mide no en función de la actividad que se exigirá al trabajador, sino de una suerte de actividad analógica, preparatoria, lúdica y casi ficticia. Aquello que miden los exámenes con relación a la demanda profesional no es lo que se esperará que haga la gente, sino más bien el nivel en que será socialmente exigible pagarla" (pág. 7).

Estas observaciones no poseen un valor absoluto, por cierto; corresponden, no obstante, a una realidad irrefutable. Por ejemplo, nunca se atribuye el grado de ingeniero a un obrero o a un técnico, por más que haya adquirido la misma competencia funcional.

El último examen que rinde con éxito confina, pues, muchas veces a una persona en una zona socioprofesional de la cual le será muy difícil salir, incluso en la actualidad. La

²² J. C. Passeron, "Sociologie des Examens", en *Education et Gestion*, 1970, 2, 6-16.

precocidad de tal fijación se halla en profunda contradicción con toda la política educativa contemporánea y, por consiguiente, de promoción permanente.

Que el examen confiere más bien un certificado social que una garantía de competencia técnica se confirmaría, además, por el hecho de que, en muchas empresas, dos personas ocupan el mismo puesto:

“Una oficialmente, porque posee todos los títulos escolares exigidos para ocuparlo (con la remuneración, el prestigio y el poder), y la otra efectivamente, porque se encarga de la parte técnica de la tarea . . . es decir, lleva a cabo “todo el trabajo” . . . A veces, la segunda, desde un punto de vista técnico, convendría mucho más para el puesto que el titular. Pero esto lo impide, precisamente, el efecto del certificado producido por el examen que, de una vez por todas y para toda la vida, tiende a definir el valor del desempeño profesional de un individuo en función de su cotización en el mercado de los títulos escolares.”

Una vez más, esta afirmación está sujeta a restricciones. En una empresa, la mera competencia técnica pocas veces es suficiente para los ejecutivos: la capacidad de comunicarse, de ejercer funciones de líder y de relaciones públicas, de poner en armonía el trabajo de administración con el de producción, etc., son otras tantas cualidades cuya manifestación escapa a menudo a quienes solo ven el aspecto técnico de las cosas. Pero esas cualidades no siempre son reales ni mucho menos. Con frecuencia se trata de un formalismo acerca del cual volveremos a hablar.

b) *Los exámenes no son socialmente neutros*

Una encuesta del INSEE de Francia revela que si bien, durante los últimos años, las posibilidades de admisión en la enseñanza superior aumentaron para todas las categorías sociales, “el aumento global de la tasa de escolarización se mantiene, aproximadamente, entre las distintas clases socio-profesionales en la misma proporción que las desigualdades anteriores. En síntesis, se trata de un simple *desplazamiento hacia arriba* de la estructura de las desigualdades” (pág. 8). El siguiente cuadro resulta ilustrativo.

<i>Categoría socioprofesional del padre</i>	<i>Probabilidades de ingreso en la enseñanza superior</i>	
	<i>1961-1962</i>	<i>1965-1966</i>
Trabajadores rurales	1,1	2,7
Agricultores	3,4	8
Obreros	1,3	3,4
Empleados	9	16,2
Ejecutivos de mediana categoría	24,9	35,4
Profesiones liberales y ejecutivos superiores	38	58,7

¿Cómo se explica ese fenómeno que, al parecer, se halla en contradicción con la escolaridad obligatoria general y cada vez más larga, el carácter gratuito de los estudios y la distribución de las aptitudes en todas las capas de la sociedad según la ley del azar?

La diferencia de la educación familiar durante los primeros años de vida engendraría rápidamente diferencias en los mecanismos intelectuales, los medios lingüísticos y las actitudes frente al trabajo: "La educación impartida por las familias predispone tanto menos al éxito en los exámenes cuanto más pertenezcan esas familias a una clase alejada de la cultura escolar, de la cultura del saber" (pág. 11).

Los exámenes "neutros" se limitarán a registrar los efectos profundos de esa primera educación. Esto resulta exacto solo en parte. Nuestra opinión ofrece mayores matices que la de Passeron; este autor no reconoce en la medida necesaria la realidad de las trabas de origen sociocultural que acabamos de mencionar. Pero coincidimos en que los exámenes no son socialmente neutros.

"Los procedimientos de calificación y los tipos de pruebas utilizados toman en cuenta, tanto como las aptitudes técnicas, por lo menos, ciertos aspectos gratuitos de la realización que no poseen ninguna importancia técnica pero que, en cambio, están fuertemente vinculados con los hábitos culturales de una determinada clase social más que con los de otra" (pág. 12). De esta manera, comprueba Passeron, los exámenes franceses

más prestigiosos *no son tanto pruebas de conocimientos como de modales, o bien, ejercicios en el uso culto del lenguaje.*

Desde este punto de vista, la disertación sigue siendo el medio preferido. Retomando la tesis sociológica de Durkheim, de Max Weber y de A. Clause, Passeron recuerda que el formalismo ha sido siempre un medio de defensa de las clases privilegiadas. En efecto, es sorprendente comprobar que se atribuye la mayor importancia a una prueba, la disertación, que es fundamentalmente la que se sustrae a toda calificación objetiva, técnica... Ahora bien, el formalismo consiste, precisamente, en definir la cultura no en función de su contenido objetivo, sino por esos imponderables que constituyen el modo, el matiz. Esa "indefinición" permite ejercer una *función de cierre* en beneficio de un grupo favorecido.

Por otra parte, la neutralidad de los exámenes científicos tampoco está mejor garantizada, porque en ellos también interviene a menudo el lenguaje, y muy pocos examinadores permanecen indiferentes ante la elegancia de la expresión, por ajena que sea al problema técnico sobre el que versa esencialmente el examen.

En cuanto a la prueba oral, no hace más que reforzar el efecto de la hermosa presentación, el lenguaje pulido, la dicción elegante.

Cabe observar, para terminar, que el florecimiento de los exámenes de índole selectiva parece característico de las condiciones culturales del siglo XIX, sobre todo del desarrollo de una burocracia muy jerarquizada al servicio de la economía capitalista. Resulta natural, en consecuencia, que la nueva forma de civilización hacia la cual se orienta nuestra evolución ponga en tela de juicio un tipo de examen concebido para otras condiciones.

11. Debilidad de numerosas experiencias docimológicas

Desde hace más de medio siglo se vienen acumulando las experiencias destinadas a demostrar los desacuerdos entre quienes califican o la falta de fidelidad de una misma persona encargada de esa tarea.

El problema existe, sin duda. Pero en muchos casos es agravado artificialmente porque se omite invitar a quienes califican a ponerse de acuerdo en cuanto a los aspectos por

considerar y a la importancia relativa que se les atribuirá. En otras palabras, se siguen registrando pacientemente los resultados de una política anárquica de exámenes.

12. Otras críticas

Muchas otras críticas se elevan aún contra los exámenes. El Informe de la Comisión Consultiva sobre los exámenes en la enseñanza secundaria de Gran Bretaña, *depositado desde 1911*, las enumera. Lo alargamos un poco, pero aun así tenemos la impresión de no ser exhaustivos.

Con respecto al alumno:

- Dedicar demasiada energía a reproducir las ideas de otros en vez de desarrollar su creatividad.
- Se lo recompensa por aprendizajes a menudo efímeros. (Quedó demostrado que hasta un 80% de los conocimientos, sobre todo los fácticos, aprendidos para el examen, han desaparecido quince meses después.)
- Adquiere una obediencia pasiva a las consignas.
- Su aptitud para expresarse prevalece a veces sobre el contenido. (Mostraremos más adelante de qué manera las deficiencias socioculturales se manifiestan sobre todo en el campo verbal.)
- Con frecuencia se abstiene de expresar un juicio personal, para acatar las ideas del profesor que toma el examen.
- Su aprendizaje puede estar viciado con un espíritu de competencia que a veces linda con el espíritu mercenario.
- Aprende a especular con su oportunidad: en el examen tradicional, el reducido número de preguntas refleja idiosincrasias del profesor y abre el camino a los "chimentos".
- Los exámenes tradicionales, su larga preparación y el período de agotamiento que sigue a esto acortan considerablemente el año escolar efectivo.
- Los exámenes tradicionales impiden el trabajo en grupo y exaltan, por el contrario, el valor de la realización individual, fuente del egoísmo.
- La amenaza del fracaso —a menudo agravada por el miedo a una sanción familiar— incita al fraude.

El éxito de los fraudulentos constituye un mal ejemplo para los alumnos honestos.

- Los exámenes ofrecen una idea falsa del trabajo adulto en el que, cuanto más complejo y difícil es un problema, tanto más se busca consejo y se recurre a las obras de referencia.

En cuanto al profesor:

- Enseña en función del examen, e incluso, de las exigencias particulares de los integrantes de una mesa examinadora externa.

Veremos que en condiciones adecuadas, el espíritu de la enseñanza y del examen ha de ser uno solo. Nos referimos aquí a exámenes ajenos a los objetivos educacionales.

- Se siente estrangulado por su metodología si el espíritu del examen es extraño a los objetivos pedagógicos que le parecen esenciales.

¿Cómo se puede poner en práctica una pedagogía del descubrimiento, de exploración personal —procesos lentos, pero generadores de aprendizajes profundos— si el examen impuesto es de carácter enciclopédico?

- Esclavo del programa, no permite a los alumnos avanzar con su ritmo propio y, por consiguiente, digerir la materia.
- Tiende a atribuir demasiada importancia a las aptitudes y conocimientos que son útiles para el examen.

Por todo lo cual se ha otorgado el nombre de *efecto de contragolpe* al fenómeno que consiste en la modificación, por parte de los profesores, del método y contenido de su enseñanza en función del desarrollo de los exámenes impuestos desde afuera.

De esta manera, importantes aspectos educativos pueden desatenderse, y muchas veces ha sido así.

En resumidas cuentas, los exámenes tradicionales suelen presentar graves defectos de construcción. Su validez es dudosa. La evaluación de los trabajos, por su parte, padece grandes deficiencias. Además, los exámenes pueden perjudicar la salud física y mental de los alumnos. Finalmente, puede agregarse que los profesores, en algunos casos, dictan sus cursos en función de un examen y no al revés. . .

Pero ha llegado el momento de otorgar la palabra a la defensa.

á. En
sulta-

enes.
es en
desde
n así

os en

eros.
mien-
, han

re el
a las
en el

onal,

u de
ario.
amen
iosin-
. y el
dera-
po y
ndivi-
edo a

para

CAPITULO 2

DEFENSA DE LA NOTA SUBJETIVA Y DEL EXAMEN

Nadie negará que se ha abusado de los exámenes hasta el punto de viciar la acción educativa de la escuela. Tampoco cabe duda de que muchos concursos fueron más bien siniestras loterías y engaños. Igualmente cierto es que a más de un puntaje se le atribuyó un significado que no tenía.

Pero parece utópico deducir que los exámenes y concursos han de proscribirse para siempre y que deben subsistir únicamente series de evaluaciones ocasionales y los informes que las sintetizan.

En la primera parte hemos visto que los distintos modos de evaluación corresponden a funciones determinadas. Basta con retomar la lista para comprobar que no todas pueden ser cumplidas por un mismo procedimiento.

Por lo demás, todos los modos de evaluación implican el establecimiento de un puntaje o de una nota que quisiéramos perfectamente objetivos o, mejor aún, perfectamente controlados: esto significa que reconocemos a los maestros el derecho de adaptar gran parte de sus evaluaciones, tanto en función del alumno como de la materia, con tal que sepan exactamente lo que están haciendo.

Antes de contemplar ciertas ventajas que muchas veces se reconocen a los exámenes, se impone, no obstante, una discusión de índole negativa.

1. La medición rigurosa tal vez resulte imposible

Indagar sobre las posibilidades de validez perfecta de los exámenes y de fidelidad de las evaluaciones depende de dos postulados señalados por J. Guillaumin¹:

- 1) "Que las producciones del alumno son por su propia índole mensurables, cuantificables."
- 2) "Que las diferencias encontradas entre las medidas aplicadas por los encargados de la calificación pueden ser reducidas."

Si estos dos postulados resultan erróneos, la docimología clásica se desmorona. Ahora bien, gente entendida en la materia ha comprobado que si bien esos grandes principios están definidos desde hace ya varios decenios, apenas si han penetrado en la escuela. Tal fracaso se debería, en buena parte, a una incompatibilidad entre la índole psicológica de la situación de enseñanza y el carácter matemático o físico de los procedimientos de medición que se proponen.

La discusión no es nueva y la solución está en el compromiso. Los considerables progresos de la psicología y de la pedagogía contemporáneas se deben ampliamente a la objetivación de la observación, es decir, a la medición. Sin embargo, los investigadores en ciencias humanas poseen en la actualidad una conciencia bastante clara acerca de sus límites y, en particular, de la imposibilidad casi general de aplicar escalas de medición matemáticamente perfectas. Saben, asimismo, que en la mayoría de los casos solo medimos comportamientos aislados de todo lo humano. ¿Cómo podrían evaluarse rigurosamente entidades hipotéticas tales como el espíritu crítico, el sentido de observación, etc., a no ser por medio de realizaciones particulares supuestamente representativas del conjunto considerado?

Así como nunca se ha conseguido, hasta ahora, medir con precisión el rendimiento de la enseñanza, por ser tan numerosos los puntos de vista que se tendrán en cuenta en distintos momentos, así tampoco puede precisarse analíticamente el

¹ J. Guillaumin, "L'aspect interpersonnel de la notation scolaire: de la docimologie à la doxologie", en *Bulletin de la Société A. Binet et T. Simon*, IV, 1968, pág. 270.

valor real de una compleja realización escolar. Aun cuando, en cálculo, el resultado solo puede estar acertado o equivocado, la trayectoria del pensamiento y el esfuerzo producido pueden variar considerablemente de una persona a otra y, por consiguiente, ser difíciles de cuantificar de una manera objetiva. En cuanto a la redacción francesa, tomada como obra de arte, escapa a la evaluación parcelaria². Por esa razón, el derecho a la subjetividad, a la reacción global, tiene, como se comprenderá, sus fervorosos defensores.

Si la respuesta a preguntas de geografía, historia, ciencias naturales, obliga al alumno a expresarse, en forma oral o escrita, ¿siempre habrá de prescindirse de la elegancia del lenguaje, el ordenamiento del pensamiento, el rigor del razonamiento?

En síntesis, a partir del momento en que se deja de basar el éxito o el fracaso en criterios rígidos y a menudo esquemáticos, las producciones de los estudiantes se sustraen a la cuantificación automática, impersonal.

Es cierto que este argumento no es suficiente para renunciar a objetivar una parte de la calificación de los alumnos; pero solo una parte. La apreciación global del maestro, con toda su agudeza, su sensibilidad, tanto en lo que se refiere a la realización material como al esfuerzo de superación y a la debilidad humana, deben conservar su lugar.

La riqueza de la enseñanza reside, ante todo, en la calidad de la relación humana que crea, y la evaluación constituye uno de los aspectos de esa relación. Si ésta se empobrece hasta el punto de permitir únicamente una comunicación impersonal, el maestro puede ser remplazado con ventaja por una máquina.

² Cabe señalar, sin embargo, que Ellis Page consiguió establecer un programa de computadora que permite evaluar en forma automática las redacciones en lengua materna. La correlación con las notas tradicionales es elevada. Con todo, debería saberse si, en ese caso, el investigador no se ha limitado, simplemente, a obtener de la computadora la imitación de los métodos imperfectos que caracterizan a los encargados de la calificación.

E. Page y D. Paulus, *The Analysis of Essays by Computer*, Washington, U.S. office of H.E.W., Project 6-1318, 1968.

2. Los maestros juzgan bien a sus alumnos

Si bien los maestros tienden a relativizar sus juicios en relación con el nivel medio del grupo, no por eso su clasificación dentro de éste deja de tener una elevada validez.

Bastaría, en consecuencia, con aplicar medidas que permitan comparar esos juicios entre las distintas escuelas (en el capítulo sobre la "moderación" veremos cómo puede lograrse), para disponer de importantes puntos de referencia.

La seguridad de juicio de los maestros se explica en virtud de distintos factores:

1. Fundan su apreciación en una prolongada y continua observación.
2. Consideran un número mucho mayor de factores (sobre todo de personalidad) que el examen.
3. En particular, pueden tener en cuenta comportamientos excepcionales (en tal ocasión, tal alumno dio pruebas de una lucidez y una originalidad poco comunes) que en el examen probablemente no aparecerán.

El valor de predicción de los juicios de los maestros (ajustados a fin de hacerlos comparables) fue puesto de relieve por una investigación de la *National Foundation for Educational Research in England and Wales*³.

En 1951 y 1952 se examinó, aproximadamente, a 1.200 alumnos de la región de Twickenham en oportunidad de su selección para el ingreso en el ciclo secundario (*Grammar Schools*). Sus resultados fueron estudiados dos y tres años después.

Pruebas realizadas al comienzo:

Tests:

1. *Test* de inteligencia verbal (V).
2. *Test* estandarizado de conocimientos de lengua materna (E1) (Preguntas de elección múltiple).
3. *Test* estandarizado de conocimientos de aritmética (A) (Cálculos y problemas).
4. *Test* de inteligencia no verbal (N/V).

³ Véase F. Yates y D. Pidgeon, *Admission to Grammar School*, op. cit. pág. 57 y sig.

5. *Test* de aptitud espacial (Sp. 1) (Espacio de dos dimensiones).
6. *Test* de aptitud espacial (Sp. 2) (Espacio de tres dimensiones).
7. *Test* de conocimientos en lengua materna (E2) (Preguntas más abiertas que E1).

Juicio de un maestro de 6° grado primario:

1. Predice el éxito en los estudios en la *Grammar School* de acuerdo con una escala de 15 grados (M).
2. Califica a sus alumnos por orden de éxito en la *Grammar School*. Los investigadores ajustan esos juicios en función de un *test* de inteligencia verbal (véase descripción del sistema en el capítulo de la "moderación") (F).

Pruebas utilizadas para medir el éxito en la enseñanza secundaria: orden al mérito conferida por el inspector de estudios, adaptada en función de los resultados obtenidos en una batería de *test* objetivos de aptitudes y conocimientos, administrados inmediatamente después de elaborarse la calificación (S.H.A.).

El siguiente cuadro muestra, para el grupo de 1951, la correlación entre cada una de las pruebas de predicción y los resultados en el transcurso de los estudios:

<i>Predictor</i>	<i>Grupo de 1951, dos años después</i>	<i>Grupo de 1951, tres años después</i>
Juicio del maestro (F)	.821	.748
(M)	.796	.722
Test verbal (V)	.789	.704
E2	.749	.623
A	.734	.659
E1	.729	.622
N/V	.648	.535
Sp.1	.565	.453
Sp.2	.491	.361

El juicio ajustado de los maestros resulta ser el mejor pronóstico. Se confirma, además, el elevado valor de pronóstico del simple *test* verbal.

Cabe insistir: esas observaciones solo son válidas en la situación considerada: el éxito en las *Grammar Schools*. No sabemos lo que sucedería con otras formas de enseñanza. Sin embargo, contamos aquí con una observación estrictamente controlada, y no tenemos razón alguna para creer que no pueda aplicarse a otros casos.

3. Validez limitada pero real de los exámenes tradicionales

Sean cuales fueren sus imperfecciones, los exámenes tradicionales no impidieron que nuestra civilización alcanzara un nivel científico nunca igualado en la historia de la humanidad. La selección producida por ellos reveló, pues, su validez; por lo menos en parte.

Es evidente: con la reserva hecha en el capítulo anterior, puede afirmarse que los exámenes tradicionales permitieron identificar a los mejores alumnos, así como también a los menos dotados, selección esencial en la forma de civilización que hemos conocido en el último centenar de años.

El hecho de que el procedimiento haya sido tosco (en el sentido de que ha pasado por alto o desperdiciado talentos, sobre todo en detrimento de las capas sociales menos favorecidas) e injusto para la población de los alumnos regulares, metida de una manera brutal en una misma bolsa, no quita nada a ese otro hecho según el cual, en una democracia poco evolucionada, los exámenes desempeñaron bien su papel.

4. Endurecerse para la vida

Según los partidarios del behaviorismo, una conducta solo se aprende si efectivamente se produce. Parece deseable, en consecuencia, que el estudiante sea llevado periódicamente a hacer un esfuerzo excepcional, a poner en juego todas sus energías, a afrontar las ordalías de los exámenes con todas sus imperfecciones e incluso sus injusticias, puesto que la vida no le ahorrará pruebas semejantes, y está bien que se encuentre preparado.

Además, la competencia, la rivalidad, son rasgos fundamentales de nuestra civilización. El examen y el concurso existen y hasta es de desear que se impongan en forma creciente ahí donde haya más candidatos que puestos o empleos disponibles, pues en caso contrario tendrían vía libre el favoritismo en todas sus formas y la ineficacia que lo acompaña.

5. Ubicarse en relación con los demás

Si una calificación desfavorable en una prueba puede causar un choque, también permite ubicarse en relación con los demás (no en lo absoluto, sino en las condiciones del examen). Quien hizo todo lo que podía, está en condiciones de ajustar de ese modo, en forma más adecuada, sus ambiciones; quien no lo hizo, tiene oportunidad de descubrir, tal vez antes de que sea tarde, las consecuencias de sus debilidades.

Por regla general, el éxito beneficia más que el fracaso, pero para algunos cierta frustración constituye un valioso aguijón.

6. Amplia síntesis e integración de los conocimientos

Es innegable que los exámenes referidos a grandes conjuntos de conocimientos obligan al estudiante a elaborar síntesis que le permiten percibir, quizá por primera vez, la economía de todo el edificio, las relaciones entre las partes y, eventualmente, los elementos comunes a diferentes disciplinas.

7. El examen externo controla al profesor

Un examen regional o nacional constituye un medio de control de la labor de los profesores.

Si bien ya no existe la posibilidad de atar a los maestros a programas superabundantes, fijados en cada detalle, no por eso ha desaparecido la necesidad de que la enseñanza se desarrolle de acuerdo con un plan de estudios preciso en sus objetivos y definido en las principales materias.

Sin duda, llegará una época en la que cada alumno trabajará de acuerdo con un programa individualizado total-

mente o a medias; pero, aun en ese caso, los fines por alcanzar tendrán que ser definidos antes de iniciarse la acción pedagógica.

Los padres y las autoridades organizadoras tendrán siempre el derecho de verificar si la misión educadora ha sido cumplida.

La existencia de un control sobre todos los elementos importantes del plan de estudios constituye, asimismo, una saludable barrera contra la laxitud y el relajamiento. Consagrar el tiempo necesario al aprendizaje es loable; haraganear resulta condenable. Esto me hace recordar a un profesor de idiomas que, en dos años, no había pasado más allá de la introducción al curso. . . Al parecer, ningún inspector advirtió este hecho.

Finalmente, el examen permite comprobar, por lo menos en cierta medida, si el profesor no concentra demasiado sus esfuerzos en algunos alumnos, en detrimento del conjunto.

8. El examen externo, "feedback" para el profesor

Tanto para nuestra salud mental como para la mayor eficacia de nuestra actuación, necesitamos estar informados acerca de la validez de lo que hacemos.

El examen bien concebido permite al profesor juzgar el valor de ciertos aspectos de su enseñanza a través de los aprendizajes realizados por los alumnos.

Sería un error evaluar a los maestros únicamente según los resultados obtenidos a corto plazo. Sin embargo, ese tipo de evaluación es el único que, aparentemente, puede realizarse en la actualidad con algún rigor científico. Ahora bien, todos aquellos que ejercen una profesión vinculada con las ciencias humanas sufren, en realidad en muy diversos grados, la falta de información segura sobre la pertinencia de su acción. Se trata de una preocupación que incumbe tanto al magistrado como al asistente social, al profesor como al sacerdote.

¿Cuándo puede afirmar con certeza un maestro que ha cumplido bien con su misión?

* * *

En resumidas cuentas, no somos partidarios de suprimir por completo los exámenes, ni tampoco de renunciar entera-

mente a la calificación subjetiva. Lo importante es adoptar una actitud que tome de cada procedimiento lo mejor y más seguro. Veremos en qué forma podrá lograrse ese objetivo.

ptar
más

TERCERA PARTE

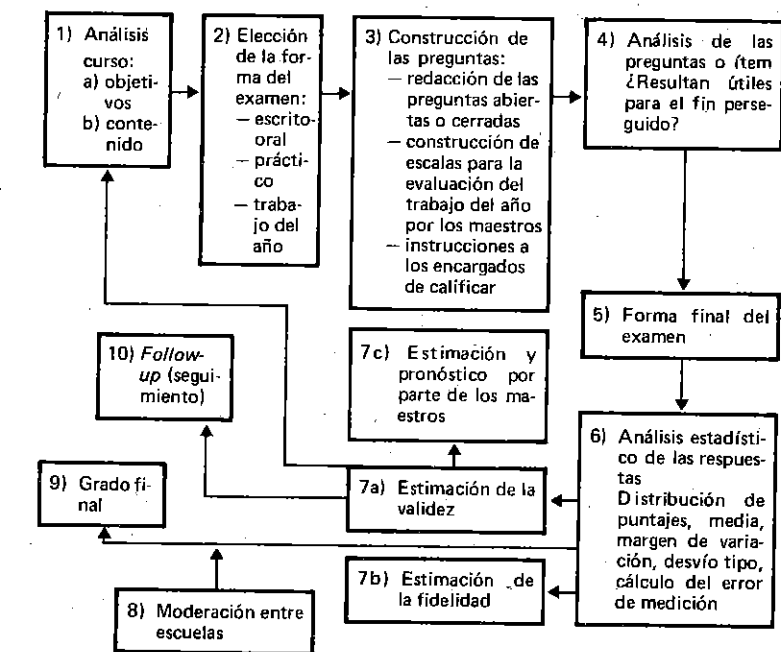
LA CONSTRUCCION DEL EXAMEN

Tal como ocurre con los *tests* estandarizados, los exámenes preparados por los maestros deberían construirse por etapas bien definidas. No es posible desplegar, ciertamente, para un simple interrogatorio de rutina, los mismos esfuerzos que para una prueba destinada a centenares de alumnos. No obstante, las exigencias de base son las mismas.

Las grandes etapas de la construcción de un examen son:

1. Definir el objeto y los objetivos.
2. Redactar las preguntas.
3. Normalizar la presentación, la administración y la corrección.
4. Estandarizar.
5. Poner a prueba la fidelidad del examen.
6. Establecer la validez.

El siguiente esquema muestra el camino que vamos a recorrer*.



*Adaptado según *Examinations Bulletin* n° 3, 1964, pág. 33.

káme-
por
ente,
erzos
s. No
son:

rrec-

os a

as
m
es
e-

1

1

CAPITULO 1

EL OBJETO Y LOS OBJETIVOS

Ver con claridad la razón de ser del examen que se prepara y definir, con la mayor precisión posible, los tipos de aprendizaje por evaluar constituyen, probablemente, la etapa más decisiva en la construcción de una prueba. ¡No solo su validez depende de ello, sino también el modo de evaluación por adoptar!

I. El objeto

Son tres los objetos posibles: el *pronóstico*, el *inventario* y el *diagnóstico*. En contados casos una prueba escolar puede incluirse exclusivamente en una de esas tres categorías. Casi siempre, los exámenes aspiran a establecer un balance, un inventario de lo adquirido, pero, al mismo tiempo, cumplen una función de pronóstico a corto o mediano plazo (¿está el alumno en condiciones de iniciar un nuevo capítulo del curso? ; ¿podrá seguir con éxito las materias del año siguiente?) y una función de diagnóstico (¿dónde tropezará el alumno?).

Por lo demás, es importante saber si el examen entra en la acción educativa de todos los días —en ese caso el alumno será comparado consigo mismo y con el grupo que lo rodea (casi siempre, su división)— o si está destinado a informar al individuo acerca de su valor relativo en su comunidad o en su país, o bien a clasificarlo en un concurso y en estos casos deberán utilizarse normas adecuadas.

A. El pronóstico

A largo plazo (más de un año, por lo menos), el pronóstico formulado sobre la base de los resultados de un examen escolar resulta a menudo decepcionante. Los cambios de maestro, de método, de espíritu del curso, pueden provocar una profunda transformación de actitud en el alumno. Además, sobre todo en los jóvenes, los intereses manifiestos sufren importantes fluctuaciones. Ha de tenerse en cuenta, asimismo, la evolución psicológica de los alumnos: por ejemplo, los profesores de enseñanza secundaria conocen bien el efecto de la crisis de la adolescencia sobre el rendimiento escolar.

Más de una vez se ha demostrado que, para un pronóstico a largo plazo, los *tests* de inteligencia son por lo menos tan buenos "predictores" como los *tests* de conocimientos y mucho más seguros que los resultados escolares¹. Esos *tests* exploran aptitudes que, sobre todo después de los 8 o 9 años de edad, suelen mostrar una gran estabilidad²: investigan el razonamiento, los factores verbal, espacial, numérico, etcétera.

Además, si se determinan las dominantes profundas de la personalidad, incluidas las zonas del interés (no de los intereses particulares, que se manifiestan en forma ocasional), la predicción del éxito escolar y universitario puede alcanzar un elevado nivel de seguridad.

A *corto plazo*, el pronóstico pedagógico se establece de tres maneras:

1) *Tests de madurez específica (readiness)*

Se refieren a las formas de razonamiento, a las aptitudes que intervienen en un determinado aprendizaje: el de la lectura, por ejemplo. La construcción, y a menudo también la utilización, de esos complejos instrumentos están reservadas al especialista. No obstante, algunos *tests* que permiten un primer reconocimiento fueron concebidos para el uso de los

¹ Véanse F. Hotyat (1962), W. McClelland (1945), E. Peel y D. Rutter (1951), P. Vernon (1957), F. Bacher (1965) y otros.

² Sin embargo, algunas modificaciones pueden todavía producirse. Sabemos en la actualidad que el cociente intelectual no es inalterable.

maestros, por ejemplo: el *Test de Inizan* para el aprendizaje de la lectura.

En los países de habla francesa sería deseable que los centros de investigación aumenten en forma considerable sus esfuerzos en el campo de la madurez específica. Ya hemos mostrado, por lo demás, nuestra escasa preparación para la mayoría de las materias y en casi todos los niveles³.

2) *Verificación de conocimientos clave o de nociones críticas imprescindibles para las nuevas adquisiciones*

Pruebas de esa índole pueden ser construidas con bastante facilidad por los maestros en cuanto tengan preparada sistemáticamente la lista de los errores cometidos en oportunidad de una enseñanza anterior⁴.

También en ese caso, fichas de investigación que precisen los requisitos previos para abordar los puntos importantes de la materia prestarían inestimables servicios a los maestros.

3) *Ensayo*

Con el propósito de determinar si un alumno está preparado para abordar una materia nueva, se le presentan sus primeros elementos y se observan sistemáticamente las reacciones. Una lección programada de tipo mixto (Skinner-Crowder) constituye, muchas veces, una feliz combinación de enseñanza y control analítico que permite un pronóstico a corto plazo.

B. El Inventario (pruebas de rendimiento)

Su finalidad primordial consiste en medir lo aprendido durante un período más o menos largo. Excepto en los casos de concursos o de *surveys* (encuestas, estudios) que preparan reformas, los inventarios cumplen en la escuela, asimismo, una función de pronóstico y diagnóstico.

En efecto, parece contrario a toda sana actitud educativa que un maestro, al comprobar graves deficiencias, no trate de

³ Véase G. de Landsheere, *Les tests de connaissances*, Bruselas, Editest, 1965.

⁴ Cada noción crítica debe figurar por lo menos en tres ejercicios; ya volveremos sobre esto.

encontrar la causa para ponerles remedio y no se preocupe por las probables dificultades que aquellas deficiencias anuncian con respecto a los aprendizajes ulteriores.

Por esta razón, sobre todo, es importante distinguir cuidadosamente en todo inventario las nociones críticas, indispensables para ciertos aprendizajes posteriores, de las nociones marginales.

C. El diagnóstico

Las pruebas de diagnóstico tratan de descubrir y explicar las deficiencias y los malos hábitos en todos los dominios del aprendizaje escolar.

Sin aventurarnos en la psicología del fracaso (que incumbe, en algunos casos, al psicoanalista), señalaremos, junto con D. Durrell⁵, que las causas más comunes de las dificultades escolares son:

1. La falta de aptitudes para realizar una tarea.
2. El conocimiento imperfecto de los elementos básicos.
3. Una enseñanza mal adaptada al nivel de aptitudes del niño y a su ritmo de aprendizaje.
4. La adquisición de malos hábitos que frenan el progreso.
5. La incapacidad para transferir, para descubrir el "sistema", para generalizar las técnicas de solución.
6. La falta de vigor al atacar los problemas, resultante de repetidos fracasos y de la ausencia de interés.

Cuando surgen dificultades escolares, el estado físico y la personalidad requieren, asimismo, la mayor atención. No olvidaremos, sin embargo, que muchos problemas emocionales del alumno provienen del fracaso y no son su causa.

En la práctica cotidiana de la escuela, el diagnóstico preciso debería ser una de las preocupaciones predominantes de los profesores y, por consiguiente, ocupar un tiempo considerable. Contrariamente a lo que una especie de mito respecto de la curva de Gauss (véase pág. 95) ha hecho creer a veces, la enseñanza ideal sería aquella que, por lo menos en

⁵ D. Durrell, *Analysis of Reading Difficulties*, New York, Harcourt, 1955.

cuanto a las nociones clave, condujese a todos los alumnos de un grupo a una ejecución perfecta.

Algunos buenos *tests* diagnósticos han sido publicados y los maestros no pueden ignorarlos. Pero en la mayoría de los casos, ellos mismos podrían construir instrumentos analíticos simples y eficaces, especialmente adaptados a sus necesidades.

Es fácil imaginar la utilidad de una planilla donde, para cada uno de los alumnos de una división, se marque con un + o un - si la respuesta fue correcta o no en tres ejercicios referidos a uno de los puntos de la materia⁶. Basta con señalar en rojo los casilleros que contienen dos o tres signos de menos para obtener una primera impresión de la situación⁷. Hace algunos años confeccionamos y llevamos una planilla de ese tipo, con la colaboración de los alumnos, referida a la pronunciación de un idioma extranjero. Los resultados fueron positivos.

Para que la prueba pueda desempeñar su rol de diagnóstico, es necesario, evidentemente, que los ejercicios se refieran a un aspecto muy específico de la materia y que se conciban, si es posible, de manera tal que no pongan en tela de juicio más de una dificultad por vez.

Como la elaboración de semejantes controles llega a convertirse fácilmente en una pesada tarea, resulta a menudo deseable la colaboración de varios profesores.

II. Los Objetivos

Es curioso comprobar que, cuando a las personas encargadas de educar a los niños se les pide que precisen los fines que persiguen, a menudo se las deja sumidas en la perplejidad.

P. Osterrieth, *Faire des adultes*, pág. 9.

⁶ En una prueba de diagnóstico se verifica, por lo menos tres veces, el conocimiento de la misma noción en distintos lugares del *test*, con el propósito de reducir en gran medida la eventual influencia de una distracción momentánea, de una respuesta correcta hallada por casualidad, etcétera.

⁷ Un ejemplo de aplicación de ese sistema se encuentra en Bongrain y otros, *Epreuves analytiques d'arithmétique* (fin de la primaria), Morlanwelz, Instituto Superior de Pedagogía, 1961.

Afirmar que los maestros no solo deben *instruir*, sino además *educar* se ha convertido en perogrullada. Sin embargo, es importante no conformarse con deseos piadosos, con vagas preocupaciones. Enseñamos, ciertamente, lo que somos: nuestra personalidad, nuestro modo de pensar y de actuar influyen de manera directa, y casi a pesar de nosotros, en los alumnos.

Pero si queremos sistematizar nuestra acción, se impone una definición exacta de los fines perseguidos.

Educar significa poner en práctica los medios apropiados para desarrollar la inteligencia y la personalidad tal como lo quiere el medio cultural, sin olvidar que la salud física exige también la atención del educador.

Instruir significa proporcionar conocimientos nuevos.

Los *objetivos generales* pertenecen a la educación; los *especiales*, a la instrucción. Pero los dos son inseparables o, por lo menos, deberían serlo.

Antes de comenzar el año lectivo, el profesor debe preguntarse acerca de los objetivos que desea lograr (en caso contrario procede a ciegas) y, *por lo tanto, definir también la materia de los exámenes*.

Sin duda, los planes de estudios y los programas escolares oficiales traen parte de las respuestas. Pero está claro que éstas solo aparecerán después de un estudio minucioso y de una larga meditación sobre esos documentos básicos. Además no es raro, sobre todo en la enseñanza secundaria, que cada materia sea objeto de un programa aparte y que únicamente los objetivos privilegiados resulten tratados en forma explícita. Incumbe entonces al profesor agregar los objetivos más generales que persigue junto con sus colegas de las otras disciplinas. Si el maestro no tiene plena conciencia de los objetivos y no se identifica con ellos, cuenta con pocas posibilidades de alcanzarlos.

Confeccionar un cuadro de los objetivos por lograr ayuda mucho al profesor en el momento de preparar su materia del año. Ese cuadro (véase el ejemplo de la pág. 65) contendrá tantas columnas como objetivos generales. Para cada capítulo o cada etapa del programa se indican los objetivos que se desea alcanzar y la relativa importancia que se les atribuye (escala de 3 a 5 grados).

La apreciación de la importancia relativa de los distintos puntos del programa escapa de la cuantificación rigurosa: ¿se

trata de una materia fundamental, de conocimiento imprescindible para abordar otras? ¿Se trata de conocimientos o capacidades indispensables para el ejercicio de una profesión o para entrar en la vida?

Por subjetiva que resulte a veces la evaluación de la importancia, el esfuerzo de reflexión que exige produce, casi siempre, una útil clarificación.

A. LOS OBJETIVOS GENERALES

En el fondo, el objetivo de la enseñanza es uno solo: formar un adulto. P. A. Osterrieth consagró páginas profundas a ese tema⁸:

Más específicamente, los maestros no deben, por una parte, perjudicar la salud física y mental de sus alumnos y, en lo posible, ayudar a desarrollarla, y por la otra, perseguir sistemáticamente objetivos cognoscitivos y afectivos que trataremos de precisar.

Separar los terrenos cognoscitivo (pensar), afectivo (estar satisfecho o irritado; amar o rechazar) y "conativo" (querer, desear) es simple teoría. ¿Podemos pensar sin experimentar sentimientos, obrar sin pensar? Es bien sabido que el organismo en su totalidad responde siempre a un estímulo, y con respecto a todo comportamiento, un análisis sutil debería revelar la importancia relativa de los componentes cognoscitivo, afectivo y psicomotor.

Nuestro propósito es mucho más modesto. Estamos buscando en la actualidad puntos de referencia que nos servirán para jalonar la enseñanza y los exámenes. Por esa razón separaremos arbitrariamente el dominio cognoscitivo del afectivo, y otorgaremos a esta última palabra una acepción muy amplia.

⁸ P. A. Osterrieth, *Faire des adultes*, Bruselas, Dessart, 1964. Véase especialmente el primer capítulo: "Le but de l'éducation" (La finalidad de la educación).

1. Los objetivos cognoscitivos

Debemos a dos investigadores estadounidenses, B. Bloom⁹ y J. P. Guilford¹⁰, clasificaciones jerarquizadas de los objetivos cognoscitivos que, pese a ciertas deficiencias¹¹, constituyen instrumentos útiles.

a) La taxonomía de Bloom

Aquí tenemos primeramente las grandes subdivisiones (en la página 67 proponemos un ejemplo de aplicación práctica):

1. Saber de memoria.
2. Comprender.
3. Aplicar.
4. Analizar.
5. Sintetizar.
6. Evaluar.

Cada uno de esos escalones ha de aceptarse en el sentido que le otorgan los autores de la taxonomía¹². Por lo demás, si bien saber de memoria y evaluar aparecen como los comportamientos cognoscitivos menos noble y más estimable, respectivamente, esto no implica, de manera alguna, que la memorización deba desaparecer de nuestra enseñanza. Por lo menos teóricamente, hay que haber pasado por cada uno de los escalones para llegar al nivel superior. La taxonomía se ha

⁹ B. Bloom y colab., *Taxonomía de los objetivos de la educación* (Dominio cognoscitivo y dominio efectivo), Buenos Aires, ed. "El Ateneo", 1971.

¹⁰ J. P. Guilford, *Modèle tridimensionnel de l'intellect*. Unas cuarenta monografías sobre el tema fueron publicadas por Guilford y su equipo (University of South California, Los Angeles).

¹¹ Véase especialmente la crítica de J. Carroll.

¹² Por ejemplo, aplicar requiere un análisis más o menos detallado. En consecuencia, ¿no habría que ubicar el análisis antes de la aplicación? Bloom y sus colaboradores limitaron el significado de *aplicación* para eludir esa dificultad. Indican, además, en varios lugares de su obra, las superposiciones de ciertos escalones y analizan, con gran sutileza, las dificultades de empleo de la *Taxonomía*. Nuestro propósito no consiste en restituir todos los matices, sino en llamar la atención sobre el instrumento.

creado para *ayudar al educador a no olvidar* ciertas etapas y para incitarlo a elevar en forma gradual el nivel de su enseñanza.

1. CONOCER DE MEMORIA

- 1.1. Datos particulares: nombres, hechos, fechas, símbolos.
- 1.2. Maneras de tratar los datos particulares (sin aplicarlos): convenciones, clasificaciones, criterios, métodos.
- 1.3. Datos universales: principios, leyes, teorías, . . .

La diferencia entre las tres subcategorías es más cuantitativa que cualitativa: saber de memoria, sin más, las fechas de nacimiento de los reyes de Francia o alguna exposición de la filosofía de Kant requiere, sobre todo, diferentes esfuerzos de memorización. En ambos casos, el nivel de pensamiento es muy bajo. Ahora bien, ¿cuántas preguntas de examen, desde la escuela primaria hasta la enseñanza superior, se sitúan en ese nivel (¿Qué sabe usted de . . .?, ¿Cuáles son las cláusulas de . . .?, ¿Cómo se prepara el . . .)?

2. COMPRENDER

Se trata aquí de la comprensión en el nivel más bajo. Mediante una nueva formulación del dato, el individuo muestra que ha superado el psitaquismo, que el mensaje le significa algo. Sin embargo, todavía no se trata de aplicar, y por consiguiente de percibir, las relaciones del dato con otros materiales, con otras situaciones.

Bloom distingue dos escalones:

2.1. Traducir, trasponer .

El contenido de la comunicación se conserva sin que se modifique el orden, pero cambia la forma.

Ejemplos: parafrasear una cita, una proposición:
"Una *taxonomía* es una *clasificación*".

Expresar verbalmente símbolos matemáticos:

$A > B$ significa que A es mayor que B.

2.2. Interpretar.

Explicar o resumir una comunicación.

La interpretación implica un nuevo arreglo, una nueva visión del material. Presupone, pues, la capacidad de reconocer y de captar las principales ideas de una comunicación y de comprender sus interrelaciones. "En ese sentido, la interpretación llega a ser sinónimo de análisis y posee ciertas características de la evaluación."¹³

Ejemplo: interpretar los datos representados en forma de cuadros o gráficos y extraer deducciones que tengan en cuenta las relaciones entre los datos o su significación de conjunto.

Ir más allá de los datos e informaciones ofrecidos: extensión de las tendencias, generalización.

2.3. Extrapolar.

3. APLICAR

La aplicación supone que la persona distingue los rasgos comunes a dos situaciones, a dos problemas; se ha producido, por lo tanto, una abstracción.

Ejemplo: empleo de procedimientos experimentales para resolver problemas domésticos.

4. ANALIZAR

4.1. Investigar elementos.

Ejemplo: distinguir los hechos de las hipótesis en una comunicación.

4.2. Investigar relaciones.

Ejemplo: ¿son lógicas las hipótesis en relación con las informaciones disponibles?

4.3. Investigar principios de organización.

Ejemplo: identificar las técnicas de propaganda empleadas en unos folletos.

¹³ B. Bloom y otros., *op. cit.*, pág. 104.

5. SINTETIZAR

5.1. Producir una obra personal.

Ejemplo: narración cautivante de una experiencia vivida.

5.2. Elaborar un plan de acción que responda a determinadas exigencias.

5.3. Derivar un conjunto de relaciones abstractas. Inducir una regla.

6. EVALUAR

Juicios cualitativos o cuantitativos establecen en qué medida el material y los métodos responden a los criterios (internos o externos).

Ejemplos: descubrir los sofismas en una discusión. Apreciar un trabajo por comparación con un modelo.

b) El modelo de J. P. Guilford

Menos utilizado hasta el presente que la taxonomía de Bloom para la preparación de los exámenes, el modelo de J. P. Guilford ofrece tal vez mayores posibilidades a causa de su rigor.

He aquí, en primer término, la forma en que J. P. Guilford y R. Marrifiels definen las tres dimensiones del intelecto y sus componentes¹⁴:

LAS OPERACIONES

Son las principales actividades o procesos intelectuales; es aquello que el organismo hace a partir de la materia prima informativa, a partir de lo que discrimina.

1. Cognición.

Conciencia, aprehensión, descubrimiento o redescubrimiento, reconocimiento, comprensión de informaciones en diversas formas.

¹⁴ Véanse los *Reports from the Psychological Laboratory of the University of South California*, publicados a partir de junio de 1950.

2. Memoria.
Retención de informaciones.
3. Producción convergente.
Generación de informaciones únicas, convencionalmente aceptadas, a partir de un dato. Se respetan el uso, la costumbre, la regla.
4. Producción divergente.
Generación de distintas informaciones a partir de un mismo dato. Originalidad, creatividad.
5. Evaluación.
Toma de decisiones o formulación de juicios que concier-
nen a la exactitud, la adecuación, la conveniencia, ...
conforme a criterios, ideales, objetivos adoptados.

LOS CONTENIDOS

1. Figurativos.
Información en su forma concreta, percibida o recordada
en imágenes.
Un mínimo de organización, de estructuración, resulta
imprescindible. Inteligencia práctica.
2. Simbólicos.
Informaciones en forma de signos desprovistos de sentido
por y en sí mismos: letras, números, notas musicales.
Inteligencia teórica.
3. Semánticos.
Informaciones en forma de significados atribuidos a las
palabras.
Inteligencia verbal.
4. De comportamiento.
Informaciones, esencialmente no verbales, que intervienen
en las interacciones humanas en las que desempeña un
papel la percepción de actitudes, necesidades, deseos,
intenciones y pensamientos del otro y de uno mismo.
Inteligencia social.

LOS PRODUCTOS

Son los resultados del tratamiento de las informaciones por el organismo.

1. Unidades.
Elementos de información relativamente aislados o circunscritos.
2. Clases.
Unidades agrupadas en función de sus propiedades comunes.
3. Relaciones.
Conexiones reconocidas entre unidades.
4. Sistemas.
Agrupamientos de unidades organizadas o estructuradas; complejos de partes en interrelación o interacción.
5. Transformaciones.
Cambios producidos en informaciones o en su utilización.
6. Implicaciones.
Extrapolación de informaciones: predicción, consecuencias, antecedentes.

Cada uno de los componentes de las tres dimensiones se combina con todos los demás. De esta manera:

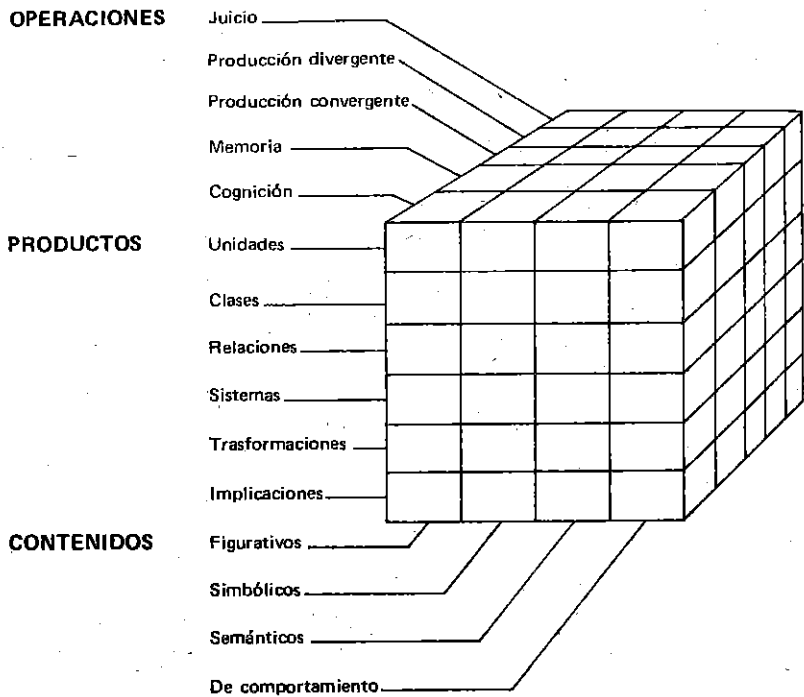
Memoria de

→ { unidades
clases
relaciones
sistemas
transformaciones
implicaciones

→ { figurativos
simbólicos
semánticos
de comportamiento

o sea, 24 combinaciones.

Como se distinguen cinco tipos de operaciones, existen en total, pues, 120 combinaciones en el modelo.



La experiencia enseña que es difícil traducir todas las combinaciones a términos utilizables para la enseñanza.

En un comienzo, los cuatro tipos de operaciones retendrán fundamentalmente la atención. De manera ideal, y desde el principio mismo de la escolaridad, el maestro debería esforzarse por introducirlos en todas las actividades.

1. La repetición de memoria demuestra la fijación de la noción.
2. La producción convergente prueba la capacidad de aplicar las adquisiciones de acuerdo con las reglas y los usos.
3. La producción divergente evidencia la capacidad de descubrir soluciones o usos nuevos.
4. Por último, la evaluación demuestra la capacidad de criticar, de expresar un juicio en función de criterios bien definidos, y de proceder de conformidad.

Con excesiva frecuencia, la actividad escolar y los exámenes que la sancionan se limitan a las dos primeras operaciones.

Esforzarse por mantener un equilibrio entre las cuatro categorías de los contenidos es importante, tanto para la enseñanza general como para la técnica o artística.

2. El dominio afectivo

Si esquematizamos el análisis detallado de P. Osterrieth¹⁵, diremos que un individuo ha alcanzado plenamente la condición de adulto:

- 1° a) Si su comportamiento ha encontrado su lógica, su coherencia y carece de versatilidad.
- b) Si ha adquirido, por lo demás, una sana tolerancia al cambio, a la contradicción, al fracaso.
- 2° a) Si ha conquistado su independencia, su autonomía intelectual y afectiva.
- b) Si, con todo, es capaz de darse, de permanecer fiel a sus compromisos y sentimientos.

La educación es un largo camino hacia este último equilibrio.

Ofrecemos a continuación una adaptación e interpretación de la taxonomía de los objetivos de D. Kratwohl y B. Bloom¹⁶. Obtenemos nuevamente cinco escalones, cinco etapas que conducen del comportamiento más pasivo al más activo.

1. El individuo responde a un estímulo exterior

1. *Es simplemente receptivo*

Constituye una especie de estado afectivo amorfo en el que la persona percibe la belleza o la fealdad, los distintos sentimientos, sin reaccionar; algo así como un espejo que no reflejará la imagen.

¹⁵ P. Osterrieth, *Faire des adultes*, Bruselas, Dessart, 1964.

¹⁶ B. Bloom y colab., *Taxonomía de los objetivos de la educación* (Dominio cognoscitivo y dominio afectivo), Buenos Aires, ed. "El Ateneo", 1971.

Por lo demás, ese comportamiento resulta difícil de distinguir de la simple cognición que precede a la incorporación en la memoria. Solo se observa cierto despertar de la atención. Ejemplo: el individuo escucha música, escucha hablar a los demás.

2. *Recibe y reacciona*

El individuo reacciona con claridad, ya sea obedeciendo o bien manifestando placer mediante la palabra, el gesto o la actitud. En ese estado todavía no se observa el rechazo explícito que daría prueba de una elección deliberada.

Para el profesor de literatura, éste es el momento en que los alumnos aún no tienen el gusto suficientemente formado como para hacer una elección personal, en que su sensibilidad aún no está bastante refinada como para permitirles aventurarse solos en el descubrimiento, pero es cuando, frente a las obras bellas, empiezan a sentir su grandeza.

3. *Recibe y reacciona aceptando o rechazando*

En este momento el individuo sabe lo que desea o lo que quiere, siempre que se le ponga frente a las personas o las cosas; se compromete.

II. El individuo toma la iniciativa

4. *Procura espontáneamente comprender, juzgar, sentir*

La persona experimenta suficiente interés y curiosidad como para instruirse sin ser invitado a ello, bastante sensibilidad como para tomar una iniciativa emocional, o ha descubierto, asimismo, el sentido de los valores, al punto de abrazar una filosofía o una religión.

5. *Procede según sus opciones*

Se trata del estado psicológicamente adulto, según la definición de P. Osterrieth.

Por ejemplo, el individuo vive en función de sus opciones morales, sentimentales, estéticas, pero también es capaz de cambiar de conducta a la luz de pruebas y argumentos convincentes.

Esta última etapa de la evolución afectiva corresponde a la evaluación en el terreno cognoscitivo.

B. LOS OBJETIVOS ESPECIALES

Teóricamente, todas las materias, todos los puntos de los programas escolares, ofrecen la oportunidad de acercarse a los objetivos generales y verificar si han sido alcanzados.

Sin embargo, se imponen tres observaciones:

1. La práctica muestra que una misma prueba solo puede referirse a un número limitado de objetivos.
2. Ciertas materias se prestan mejor que otras para alcanzar determinados objetivos.
3. El grave problema de la trasferencia de lo aprendido sigue planteándose en forma aguda. Por ejemplo, cultivar la divergencia en cuanto a las actividades artísticas no garantiza, de manera alguna, que ese rasgo se manifestará en los terrenos científicos o en la vida práctica.

Sea como fuere, el primer paso hacia la identificación de los objetivos especiales es siempre fundamentalmente el mismo y consiste en elaborar una planilla con doble entrada. En el extremo superior se anotan los objetivos generales; al costado se inscriben las materias del curso. Cada intersección de columna y renglón indica, en principio, un objetivo especial.

EJEMPLO: QUIMICA¹⁷

	<i>Conocimiento de hechos, métodos y técnicas</i>	<i>Apli- cación</i>	<i>Evalua- ción</i>
Equilibrios iónicos 1. Generalidades 1. Grado de ionización o fracción ionizada 2. Mediciones del grado de ionización 3. Aplicación de la ley de acción de las masas sobre los equilibrios iónicos 4. Ley de la dilución de Ostwald 2. Producto de solubilidad 1. Definición 2. Cálculo del producto de solubilidad en función de la solubilidad 3. Disminución de la solubilidad de un electrólito poco soluble			

¹⁷ Fuente: Berger y Dighaye, *Chimie IV*, Lieja, Sciences et Lettres, 1967.

	Conocimiento de hechos, métodos y técnicas	Aplicación	Evaluación
<p>4. Precipitación de un electrólito por agregarse a su solución saturada un electrólito que aporta un ion común</p> <p>5. Disolución de los precipitados</p> <p><i>Apéndice:</i> Aplicación del producto de solubilidad a la precipitación de los sulfuros</p> <p>3. Producto iónico del agua y pH</p> <ol style="list-style-type: none"> 1. Producto iónico del agua 2. El pH <ol style="list-style-type: none"> a. La notación pH b. pH del agua pura c. La escala de los pH d. Determinación del pH e. Los indicadores coloreados f. Métodos experimentales de determinación del pH g. Distinción entre electrólitos (ácidos y bases) fuertes y débiles por determinación del pH de soluciones diluidas de concentraciones conocidas h. Cálculo del pH de las soluciones de ácidos y bases a partir de sus concentraciones molares y, eventualmente, de sus constantes de ionización <p>4. Soluciones tampón (<i>buffer</i>)</p> <ol style="list-style-type: none"> 1. Introducción experimental 2. Cálculo del pH de una mezcla tampón <p>5. Hidrólisis</p> <ol style="list-style-type: none"> 1. Generalidades 2. Caso de sales derivadas de un ácido y una base fuertes 3. Caso de sales derivadas de un ácido fuerte y una base débil 4. Caso de sales derivadas de un ácido débil y una base fuerte 5. Caso de sales derivadas de un ácido y una base débiles 6. Factores que influyen en la hidrólisis <p>6. Neutralización de ácidos y bases</p> <ol style="list-style-type: none"> 1. Calor de neutralización 2. Curvas de neutralización <ol style="list-style-type: none"> a. Acido fuerte - base fuerte b. Acido débil - base fuerte c. Base débil - ácido fuerte <p>7. Los métodos de análisis cuantitativo por vía química</p> <ol style="list-style-type: none"> 1. Métodos gravimétricos 2. Métodos volumétricos <ol style="list-style-type: none"> Ácidos - bases Elección de los indicadores Titulaciones <i>redox</i> (de oxidorreducción mutua) 			

La materia "Química", que figura en la planilla anterior, constituye uno de los cuatro capítulos de un curso destinado al año superior de la enseñanza secundaria. Solo expusimos tres de los numerosos objetivos generales que son posibles.

Una simple ojeada sobre el conjunto revela que, aun en esas condiciones muy simplificadas, el profesor se halla ante una elección difícil. ¿Sobre qué puntos concentrará el examen a fin de obtener un muestrario suficiente de la materia? ¿Qué importancia relativa asignará a las materias elegidas y a los distintos objetivos especiales que les corresponden? ¿En qué consiste exactamente la evaluación en un curso como éste?

Según vemos, el problema no es sencillo ni mucho menos. Supera casi siempre la competencia de un solo hombre y desemboca en cuestiones que han quedado sin respuesta. Convendría crear comisiones de definiciones de objetivos en las que docentes, psicólogos y especialistas en investigación educativa unan sus esfuerzos.

Presentamos ahora una serie de ítem de ciencias que ejemplifican los principales niveles de la jerarquía de B. Bloom.

TAXONOMIA DE BLOOM

Ejemplo: Ciencias naturales¹⁸

1. Conocimiento

a) Conocimiento de datos específicos.

Los hidratos de carbono y los lípidos son utilizados por el organismo:

- A. Como alimentos energéticos.
- B. Para formar nuevo citoplasma.
- C. Como fuente de sales minerales.
- D. Como alimentos plásticos.
- E. Como fuente de vitaminas.

b) Conocimiento de la forma de tratar informaciones específicas.

¹⁸ Adaptado según D. G. Lewis, "Objectives in the Teaching of Science", en *Educational Research*, jun. 1965, págs. 186-199.

Equilibre la ecuación: $\text{H}_2\text{S} + \text{O}_2 \rightarrow \text{H}_2\text{O} + \text{SO}_2$
y cuente el número de moléculas del conjunto.

- A. 4
- B. 18
- C. 5
- D. 9
- E. Otro número distinto de los anteriores

c) Conocimiento de los principios generales.

Una máquina simple no puede:

- A. Cambiar la dirección de una fuerza aplicada.
- B. Sustituir una fuerza pequeña por otra grande.
- C. Ofrecer una ganancia de 100 entre las fuerzas puestas en juego.
- D. Vencer una resistencia mayor que la fuerza aplicada.
- E. Aumentar la energía del sistema.

2. *Comprensión* (el alumno conoce y sabe utilizar la materia que le fue transmitida sin que estuviera en condiciones, como es necesario, de establecer relaciones con otras materias).

Si un niño, en épocas de mucho frío, pone la mano sobre un poste de acero, puede que ella quede pegada a éste, lo que no sucedería si el poste fuese de madera.

Esto ocurre porque el acero:

- A. Tiene un punto de fusión más elevado que la madera.
- B. Irradia el calor más que la madera.
- C. Posee un peso específico más elevado que la madera.
- D. Es mejor conductor de calor que la madera.
- E. Posee propiedades magnéticas que la madera no tiene.

3. *Aplicación* (abstraer, de una situación particular, y aplicar la abstracción de otra manera).

Cuando el magnesio arde para formar óxido de magnesio, MgO , dos gramos de oxígeno se combinan con 3 gramos de magnesio. Si se quemaran 6 gramos de magnesio en un recipiente cerrado que contiene 3 gramos de oxígeno:

- A. No habrá reacción.
- B. Se formará Mg_2O .
- C. Se formará MgO_2 .
- D. Quedará cierta cantidad de magnesio.
- E. El óxido de magnesio contendrá un menor porcentaje de oxígeno.

4. Análisis.

Si dos objetos, X e Y, tienen el mismo momento, la energía cinética de Y sólo puede ser superior a la de X si:

- A. Y es más pesado que X.
- B. Y se desplaza más rápidamente que X.
- C. Y tiene el mismo peso que X.
- D. Y se desplaza más lentamente que X.
- E. Y está situado a una distancia mayor del eje que X.

5. Síntesis (combinar los elementos para formar un todo estructurado)¹⁹.

Algunos problemas que llamaron la atención de los hombres de ciencia pueden incluirse en el rubro "estructura de la materia". En ese terreno, las primeras investigaciones se referían a la estructura discontinua de la materia, hipótesis emitida por Dalton en 1805.

Otro grupo de investigaciones se relaciona con la teoría cinética de los gases. Apareció en el siglo XVII y fue retomada de manera más completa por Maxwell (1859) y Boltzmann (1896).

Muestre de qué manera la teoría cinética de los gases constituyó un argumento a favor de la hipótesis de la representación atomística de la materia.

Su exposición debe abarcar:

1. Un resumen de las hipótesis fundamentales de esa teoría.
2. Algunas deducciones importantes de esa teoría, aplicada al caso de los gases perfectos.
3. La interpretación de las leyes experimentales de Gay-Lussac y de la hipótesis de Avogadro.

¹⁹ Este ítem pertenece a H. Guillaume.

6. *Evaluación* (juicios acerca del valor de los argumentos que sirven para una demostración, o bien, acerca de los criterios).

La teoría atómica de Dalton es:

- A. Un buen ejemplo de verdad absoluta.
- B. Un buen ejemplo de una concepción estática de la ciencia.
- C. Exacta y permitió hacer otros descubrimientos científicos.
- D. Parcialmente falsa, pero permitió hacer otros descubrimientos científicos.
- E. Ni correcta ni equivocada y fue dejada de lado.

Por último, aquí tenemos algunos ejemplos de preguntas referidas a tres aspectos característicos del modelo de J. P. Guilford:

<i>Materia</i>	<i>Producción convergente</i>	<i>Producción divergente</i>	<i>Evaluación</i>
Ciencias	Explique por qué no puede haber vida en Mercurio.	¿En qué podría diferir la vida en Marte de la nuestra?	¿Cree usted que hay vida en Marte?
Geografía	¿Qué influencia ha tenido el estrecho de Bering sobre la población de América del Norte?	¿Qué habría sucedido en caso de no haber existido el estrecho de Bering?	¿Cuál es, en su opinión, el papel principal que desempeña en la actualidad el estrecho de Bering?
Historia	Explique la importancia del descubrimiento de América por Colón para la vida económica de Europa.	¿Qué habría sucedido si Colón hubiera descubierto el camino a la India y no el de América?	¿Cuáles son, en su opinión, las dos consecuencias más importantes del viaje de Colón?
Lengua materna	Explique por qué la novela se ha desarrollado mucho más en EE.UU. que en Europa.	He aquí el comienzo de una novela. Imagine tantos desenlaces como pueda.	¿Qué es más importante en una novela: los caracteres o la trama?

REDACCION DE LAS PREGUNTAS

I. Observaciones generales

Para que un examen resulte equitativo y válido debe contener casi siempre gran cantidad de preguntas. Si el terreno por cubrir es extenso, esto parece evidente. Preguntar tan solo acerca de una fracción, a veces mínima, de la materia, presupone una exigencia difícil de justificar, sobre todo en nuestro actual sistema de enseñanza, no individualizado¹: se considera que todos los alumnos han aprendido todo y que la calidad de todo lo aprendido es homogénea. Por lo tanto, preguntas referidas arbitrariamente a una parte del todo permitirían llegar a un juicio valedero.

En realidad, si a causa de la mala suerte el profesor pregunta acerca de las pocas cosas que el alumno no estudió, por ejemplo una vigésima parte del conjunto, es una catástrofe. Además, muchas veces una simple distracción, un cansancio momentáneo, explican una respuesta errónea en un campo bien conocido.

Ya hemos visto que en los *tests*, para prevenir ese peligro, se trata por lo general de introducir tres veces como mínimo la misma pregunta en diferentes formas; esto, por cierto, no suele hacerse en los exámenes tradicionales.

Por lo demás, la multiplicidad de los objetivos, tal como nos ha revelado el modelo de Guilford, hace difícilmente concebible un número reducido de preguntas.

¹ Véase en la 5a. parte "El mito de la curva de Gauss".

Sin embargo, sería poco realista recomendar una construcción científica para cualquier pequeño ejercicio de control o de diagnóstico.

Saber en qué forma han de formularse las preguntas es probablemente la capacidad más necesaria en un profesor. Pero es un arte difícil, tal como queda demostrado en los errores que se advierten en la elección del nivel de lenguaje y en las oscuridades que tantas veces se encuentran en la redacción de los ejercicios de aplicación y en las preguntas de examen.

A. Preguntas comprensibles

Muchas respuestas erróneas no se deben al desconocimiento de la materia, sino a malentendidos, a una mala comprensión de las preguntas. El peligro aparece especialmente en un terreno como las matemáticas, porque en él se combinan fácilmente tres dificultades: la inherente al problema, la del lenguaje abstracto de las matemáticas y las deficiencias lingüísticas de los mismos matemáticos.

Pero, según sabemos, lo abstruso no perdona tampoco a los historiadores, los geógrafos, los profesores de ciencias. . . o de lenguas.

Al no cerciorarse de la claridad de las preguntas, los maestros corren el peligro, sobre todo, de cometer una injusticia social. La investigación contemporánea confirma que muchos niños provenientes de medios sociales poco favorecidos sufren de graves deficiencias en el terreno del lenguaje. Por ejemplo, Davis y Haggard mostraron que basta con modificar la forma de un problema, sin cambiarle el sentido, para que la diferencia de éxito entre los niños de un medio socioeconómico favorecido y otros que no pertenecen a él pase del 12 al 32%².

B. Tener en cuenta el nivel de información

Para verificar la capacidad de transferencia de lo aprendido, a menudo resulta indispensable presentar los problemas en un

² A. Davis, "Education for the Conservation of Human Resources", en *Progressive Education*, 1950, 27, 221-224.

contexto todavía no visto en clase. Aun así, es necesario que el contexto tenga un sentido para el alumno.

Algunos pequeños problemas, encontrados hace poco en un manual, suponían el conocimiento de las reglas de juego del tenis o de la manera de llenar las boletas de pronósticos de fútbol. ¿Qué significarán en ese caso los fracasos, si los alumnos no tienen la posibilidad de informarse acerca del sentido real de las preguntas antes de tratar de contestarlas?

C. Poner a prueba o "testear" las preguntas

En la medida de lo posible, hay que poner a prueba las preguntas antes de presentarlas en un examen. De esta manera se descubren las imprecisiones, las deficiencias de redacción, los errores materiales y, asimismo, el nivel de dificultad.

Es difícil que un solo docente pueda llevar a cabo semejante prueba; una vez más, el trabajo en grupo ofrece muchas posibilidades en ese terreno, preparándose el examen con mucha anticipación.

D. Cálculo de la facilidad de las preguntas

Si es posible "testear" con un gran número de sujetos, ese cálculo se llevará a cabo antes del examen. En caso contrario, resulta útil, sin embargo, hacer la operación, ya sea para percibir mejor la fisonomía de las respuestas, o bien para una utilización posterior.

El *porcentaje de éxitos* es el índice más sencillo. Tengamos en cuenta, no obstante, que no es correcto calcular porcentajes en función de números reducidos. Normalmente, el número de sujetos debería superar la centena.

Una vez conocido el índice de facilidad, el profesor se halla en mejores condiciones de dosificar el examen.

E. Cálculo de la eficacia - Poder discriminatorio

El índice de facilidad por sí solo puede inducir a error, porque a veces es el resultado de factores accidentales. Por esa razón, conviene combinarlo con el *índice de eficacia*, que revela en qué medida una determinada pregunta diferencia los buenos alumnos de los regulares.

1. Método simple

Pidgeon y Yates proponen un método sencillo de cálculo, muy útil para los educadores:

- Dividir la clase en tres grupos: superior, medio e inferior, sobre la base de los puntajes totales en el examen considerado.
- Ver, para cada ítem, qué porcentaje de cada grupo tuvo éxito.
- La diferencia entre el porcentaje de éxitos del grupo superior y del grupo inferior proporciona una buena estimación de la eficacia para cada ítem.

Si la diferencia es escasa o nula, la pregunta no distingue bien entre los buenos alumnos y los regulares.

Veremos más adelante que cuanto más sutil resulta la discriminación tanto más se amplía la distribución de los resultados (la curva de Gauss se aplanan).

2. Método más sutil

He aquí otro método, utilizado para el examen final del colegio secundario en Inglaterra³.

Se constituyen cuatro grupos, A, B, C, D.

Se proporcionan las siguientes consignas:

- 1) Determinar, para cada uno de los ítem y en cada uno de los cuatro grupos, el número de alumnos que responden correctamente.
- 2) Calcular la media obtenida por cada uno de los cuatro grupos en cada ítem o pregunta.
- 3) Reunir todos esos datos en un cuadro de conjunto y comparar las medias de cada ítem en los cuatro grupos.

Si las cuatro medias se ordenan de la misma manera que los puntajes medios del *test* completo en los cuatro grupos, puede afirmarse que el ítem considerado contribuye a la discriminación total del *test*. En caso contrario, el ítem es sospechoso.

³ D. Mather y colab., *The C.S.E., A Handbook for Moderators*, Londres, Collins, 1965, pág. 108.

Ejemplo:⁴

Pregunta n°	1	2	3	4	5	6	7	Total
Máximo	6	6	6	6	10	8	8	50
Grupo A	4,9	5,1	5,3	5,2	4,3	5,8	4,3	34,9
Grupo B	3,2	4,9	5,8	4,1	4,7	5,9	3,8	32,3
Grupo C	3,4	4,6	5,4	3	3,4	3,6	3,9	27,3
Grupo D	2	3,4	5	2,8	2,3	4,1	4,1	21,8
Orden de los puntajes medios	A C B D	A B C D	B C A D	A B C D	B A C D	B A D C	A C B D	A B C D
Discrimina	No	Si	No	Si	No	No	No	Si

La respuesta obtenida de esta manera carece de precisión. En cuanto a los ítem considerados como aceptables, no podemos afirmar si discriminan con precisión o, por el contrario, si son simplemente aceptables. Además, no sabemos bien cómo mejorar la redacción de la pregunta o el modo de evaluarla.

A fin de obtener informaciones más exactas, los expertos en estadística utilizan técnicas complicadas⁵. La siguiente solución gráfica es sencilla y brinda los matices deseados.

Consideremos los resultados individuales de cada uno de los 20 alumnos de cada grupo. Tomaremos como ejemplo las preguntas 2, 3 y 4. En vez de escribir los resultados en números, los marcamos con rayitas en una ficha.

⁴ A partir de un total de 80 trabajos, D. Mather forma cuatro grupos de 20: A = los 20 primeros; B = los 20 siguientes, y así con los restantes.

⁵ Aunque el uso de la computadora las pone hoy al alcance de todos.

Pregunta 2 - Máximo 6				
Punt. obt.	Grupos			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Pregunta 3 - Máximo 6				
Punt. obt.	Grupos			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Pregunta 4 - Máximo 6				
Punt. obt.	Grupos			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Si indicamos, en fichas similares, el cuartil superior, la mediana y el inferior⁶, el comportamiento de los alumnos aparece con claridad.

⁶ LA MEDIANA: la nota del medio en una serie de notas ordenadas;

Pregunta 2				
Puntos	A	B	C	D
6	QS	QS		
5	M	M	QS	QS
4	QI		M	
3		QI	QI	
2				M
1				QI
0				

Pregunta 3				
Puntos	A	B	C	D
6	QS	QS	QS	QS
5	M	M	M	M
4	QI	QI	QI	QI
3				
2				
1				
0				

Pregunta 4				
Puntos	A	B	C	D
6	QS	QS		QS
5	M	M	QS	
4	QI		M	
3		QI		M
2			QI	
1				QI
0				

si el número de notas es par, se calcula la media aritmética entre las dos medianas

2 4 8 10 11 13

- 9 -

En una escala de medidas, la mediana es el punto por encima del cual se halla exactamente la mitad de los casos.

El CUARTIL SUPERIOR o Q_s es la nota del medio de la mitad superior de la serie.

El CUARTIL INFERIOR o Q_i es la nota del medio de la mitad inferior.

En el caso ideal, deberíamos observar, para cada una de las medidas, un declive continuo de A hacia D que resultará más o menos marcado según el poder discriminatorio. Una elevación indica que un grupo inferior (según el resultado en el conjunto del examen) obtiene más puntos que el superior.

Se observa:

Pregunta 2:

— No hay ninguna elevación y se advierte una tendencia a bajar, pero no muy nítida.

Item levemente discriminatorio.

Pregunta 3:

— Ningún descenso en el cuartil superior.

— Elevación en la mediana.

— Elevación en el cuartil inferior.

Item no discriminatorio. Debe rechazarse.

Pregunta 4:

— Ninguna elevación.

— Descenso bien marcado en la mediana.

— Descenso muy marcado en el cuartil inferior.

Buena discriminación. Cabe hacer una reserva, sin embargo, en cuanto al QS del grupo D: los correctores fueron demasiado generosos con los mejores alumnos del grupo débil (D).

II. ¿Respuestas abiertas o cerradas?

A una pregunta de *respuesta abierta*, el alumno contesta espontáneamente utilizando su propio vocabulario. Hablamos de *respuesta cerrada* si el sujeto debe elegir entre varias respuestas que se le proponen.

Tradicionalmente, la escuela operaba en forma exclusiva con el primer tipo de preguntas; con esto se privaba de un instrumento de gran utilidad.

A. Respuestas abiertas (orales o escritas)

Son las preguntas más naturales, las que planteamos en cualquier momento de la vida.

Convienen especialmente para pruebas de control tomadas rápida y espontáneamente por los maestros durante la ense-

ñanza, o bien para verificar aprendizajes tan complejos que escapen al análisis riguroso.

La evaluación de las capacidades superiores (creatividad, juicio, espíritu crítico...) está, aparentemente, dentro de su campo de aplicación. Pero no en forma exclusiva. En primer término, porque, como en seguida veremos, las preguntas de elección múltiple bien redactadas permiten exploraciones mucho más sutiles de lo que parece. Luego, porque, preocupados por la objetividad y el rigor, quienes califican tienden, tal vez de manera inconsciente, a retener de las respuestas abiertas tan solo los elementos más concretos, los más fácticos. ¿La divergencia de las evaluaciones no dependen de la sutileza del objeto al cual se refieren?

“Debemos admitir —escribe Vernon⁷— que los alumnos excelentes tienen ocasión de mostrar ciertas cualidades excepcionales en el examen tradicional, y que una persona perspicaz encargada de la calificación tal vez puede percatarse de ellas, mientras que es muy factible que el resto tenga dificultades con la respuesta a causa de su anticonformismo.” El debate se halla lejos de estar cerrado.

Al brindar la respuesta enteramente por sí mismo, el alumno procura demostrar dos cosas: su conocimiento de la materia y su capacidad para expresarlo verbalmente.

Durante mucho tiempo se creyó que las dos cosas corrían paralelas. La vieja sentencia: “Lo que se comprende bien, se enuncia con claridad y las palabras para expresarlo llegan fácilmente” ha constituido una verdad para muchos maestros hasta el día de hoy. Sin embargo, las cosas no son tan sencillas. La traducción verbal del pensamiento no es sino una forma de expresión despreciada, entre otras, durante siglos por las clases sociales privilegiadas, a las cuales les repugnaba todo trabajo manual, todo trato con la realidad.

Renunciar completamente a las preguntas de respuesta abierta sería un error en una civilización en la cual sigue predominando la comunicación verbal. Pero a veces es bueno aislar, por lo menos en parte, los conocimientos y la capacidad de expresión verbal, y esto es permitido por las respuestas cerradas. Los profesores no deben vacilar en recurrir a ellas, tanto menos cuánto que, tal como señalan

⁷ Bull 4, pág. 7.

Pidgeon y Yates, nunca se ha demostrado que, aun utilizadas con frecuencia, hayan perjudicado el desarrollo verbal.

Una respuesta formulada con toda libertad presenta otro grave inconveniente: su unicidad dificulta la evaluación, por comparación con las respuestas de otros individuos. Esta es, sin duda, una de las principales fuentes de desacuerdo entre los encargados de corregir, denunciada tan a menudo por la docimología.

B. Respuestas cerradas. Preguntas (ítem) de elección múltiple

1. Utilidad

El examen intensivo aspira a verificar en detalle la calidad de lo adquirido; persigue a menudo un objetivo de diagnóstico. El examen extensivo se refiere a una materia amplia.

En ambos casos se requieren numerosas verificaciones. Por esa razón se recurre a las preguntas de "respuesta cerrada", sobre todo del tipo "de elección múltiple".

Ejemplo: el primer test de inteligencia utilizable en la práctica corriente fue construido por:

- a) Binet.
- b) Galton.
- c) Goddard.
- d) Spearman.
- e) Terman.

2. Elaborar una lista de preguntas

Encontrar varias decenas de preguntas precisas para un solo examen somete a la imaginación a una ardua prueba. Crear un fichero de ítem y enriquecerlo en toda oportunidad facilita mucho la tarea.

No se anota más de una sola pregunta en cada ficha; para poder inscribir, asimismo, los "distractores", es decir, las respuestas falsas, pero verosímiles, a medida que se van descubriendo. Esos "distractores" surgen sobre todo de los errores cometidos con frecuencia por los alumnos; de esta manera se tiene la seguridad de respetar su "lógica" y está garantizada la verosimilitud.

Los profesores de una misma asignatura pueden unir fácilmente sus esfuerzos en esa materia.

Nos encaminamos, al parecer, hacia la creación de *bancos de ítem*, de modo que centros especiales puedan poner a disposición de los maestros varios centenares de preguntas de elección múltiple, bien actualizadas para determinadas poblaciones.

3. *Explotar la gama de posibilidades lógicas*

Responder a un ítem de elección múltiple no constituye necesariamente un simple ejercicio de memoria, ni mucho menos⁸.

Las investigaciones realizadas por los servicios de enseñanza superior de Francia, en vista de la reforma de los exámenes de medicina, pusieron de relieve con toda claridad la riqueza del sistema⁹. Elegimos a propósito ejemplos tomados de los exámenes universitarios, con la esperanza de convencer a los maestros de que, sea cual fuere la materia que enseñan, el hecho de recurrir al examen objetivo no queda excluido a priori.

En el presente examen se distinguen ocho tipos de preguntas:

a) *Pregunta con complemento simple.*

El ítem se presenta, en su espíritu si no siempre en su forma, como una oración por completar:

Ej.: Entre las características siguientes, la que se aplica a todas las enzimas es:

- A. Contienen siempre una coenzima dissociable.
- B. Son termoestables.
- C. Contienen siempre nitrógeno en su molécula.
- D. Contienen siempre fósforo en su molécula.
- E. Son dializables.

⁸ La técnica ha sido aplicada con éxito incluso en el análisis literario - véase B. Choppin y A. Purves, *A comparison of open-ended and multiple choice items dealing with literary understanding*, en *Research in the Teaching of English*, 3, 1, 1969, 15-24.

⁹ París, "Ministère de l'Education nationale. Enseignement supérieur, Examens et concours, section médicale 1961". Todos los ejemplos médicos que vienen a continuación fueron tomados de esa publicación.

No siempre es necesario que la respuesta correcta figure entre las elecciones propuestas. C hubiera podido ocupar el lugar de E y convertirse en: "Ninguna de las respuestas anteriores".

El ítem puede adoptar la forma negativa:

Ej.: Un hidrosol metálico tiene todos los caracteres siguientes, menos uno. Indique cuál.

A — B — C — D — E

Para responder, el candidato tiene que conocer todas las características del hidrosol. Sin embargo, la forma negativa obliga a una gimnasia intelectual que se superpone a la dificultad inherente a la materia.

b) Asociación simple.

Sirve para verificar el conocimiento "de cierto número de entidades que pueden estar o no relacionadas".

Se trata aquí de relacionar un elemento precedido de una letra con otro precedido de un número (asociación o apareamiento).

Ej.: *Rama posterior del nervio radial.*

- A. Segmento de origen.
- B. Grupo de ramas posteriores.
- C. Grupo de ramas anteriores.
- D. Nervio interóseo posterior.
5. Está aplicado sobre la faz posterior del ligamento interóseo.
6. Rodea el cuello del radio.
7. Provoca la inervación de los músculos de la capa superficial de la región antebraquial posterior.
8. Pasa entre las dos cabezas del supinador corto.
9. Determina la inervación de los músculos de la capa profunda de la región antebraquial posterior.

c) Asociación compuesta.

Es una simple variante de la anterior.

Ejemplo:

- A. Paludismo de *plasmodium vivax*.
- B. Paludismo de *plasmodium falciparum*.

- C. Ambos a la vez (A y B).
- D. Ninguno de los dos.

Pregunta 1:

1. La evolución clínica tiene todas las posibilidades de ser menos grave en un hombre de raza negra que en uno de raza blanca (respuesta B).
2. Una combinación de primaquina y cloroquina es la terapia de elección en caso de un ataque agudo (respuesta A).
3. Los episodios clínicos quedan suprimidos mediante la ingestión de cloroquina una vez por semana en zona endémica (respuesta C).
4. Cura en forma definitiva con el tratamiento de cloroquina (respuesta B).
5. La contaminación se evita mediante la ingestión de cloroquina una vez por semana (respuesta D).

Cabe señalar que el número de preguntas puede en este caso aumentarse o disminuirse según la importancia acordada a la materia.

d) Asociación con término excluido.

En el ítem siguiente, 4 de los 5 fenómenos enumerados son comunes a una de las afecciones A, B, y C. Hay que indicar la afección (A) y el fenómeno que *no* le corresponde (2).

Ejemplo:

- A. eosinofilia de importancia diagnóstica
- B. plasmocitosis de importancia diagnóstica
- C. linfocitosis de importancia diagnóstica

1. triquinosis
2. mieloma múltiple
3. síndrome de Loeffler
4. enfermedad de Hodgkin
5. esquistosomiasis

e) Análisis de la relación de causa a efecto.

Ej.: La articulación radiocubital superior permite movimientos de rotación limitados PORQUE la cabeza del radio está rodeada por el ligamento anular.

- A. El enunciado y la razón propuesta son correctos y existe entre ambos una relación causa-efecto.
- B. El enunciado y la razón propuesta son correctos y no existe entre ellos relación de causa-efecto.
- C. El enunciado es correcto, pero la razón propuesta es errónea.
- D. El enunciado es equivocado, pero la razón propuesta es un hecho o un principio aceptado.
- E. El enunciado y la razón propuesta son erróneos.

f) **Análisis de observación.**

El complejo conjunto que viene a continuación coloca al candidato en una situación comparable a la experiencia real.

Historia clínica: El enfermo es un hombre de 21 años que se queja de malestar, tos y fiebre.

La enfermedad empezó, diez días antes de la internación, con malestar y una tos sin expectoración, seguidos a las 24 horas de una temperatura que oscilaba entre 37,8 y 38,3 y que ha persistido hasta el momento de la internación.

El cuarto día de la enfermedad la tos empeoró y produjo pequeñas cantidades de expectoración blanca y viscosa.

Tres días antes de la internación empezaron accesos paroxísticos de tos, a veces seguidos de vómitos. El enfermo experimentó sensaciones de escalofríos, pero sin temblor. Un dolor paraesternal anterior a la tos existe desde el quinto día de la enfermedad. En el examen clínico se comprueba una temperatura de 38,3, un pulso de 110, un ritmo respiratorio de 32 y una tensión máxima de 10% y mínima de 8.

El enfermo está bien desarrollado, no ha enflaquecido; la enfermedad parece aguda; existe disnea pero no cianosis. El examen clínico de la caja torácica muestra vibraciones vocales en el momento de palpar y auscultar. El murmullo vesicular es normal.

En la axila izquierda se escuchan unos leves estertores y la calidad bronquial del sonido está aumentada, si bien alcanza una intensidad normal.

La fórmula sanguínea es la siguiente: glóbulos blancos 3.400 (polinucleares 30%, linfocitos 62%, monocitos 5%, eosinófilos 3%).

La radiografía del tórax revela un aumento de la densidad de la región perihiliar, con áreas mal definidas de densidad desigual, nublosa en ambas bases así como en un campo pulmonar superior izquierdo.

Preguntas:

1. *¿Cuál es el diagnóstico más probable?*
 - a) tuberculosis
 - b) neumonía neumocócica
 - c) neumonía (primaria atípica) virósica
 - d) coccidioidomicosis
 - e) bronconeumonía
2. *¿Cuál es el síntoma físico que probablemente se agrega?*
 - a) esplenomegalia
 - b) síntoma de dolencia meníngea
 - c) ruido de fricción pleural
 - d) cambios frecuentes en la distribución de los síntomas torácicos
 - e) síntoma de condensación lobular izquierda
3. *¿Cuál de los siguientes exámenes de laboratorio coincide con el diagnóstico?*
 - a) la elevación y el aumento de las aglutininas frías
 - b) hemocultivo positivo
 - c) marcada leucocitosis al comienzo de la convalecencia
 - d) examen de las expectoraciones
 - e) dermorreacción positiva
4. *¿Qué terapia deberá emplearse?*
 - a) reposo en cama y estreptomycinina
 - b) reposo en cama y penicilina
 - c) estreptomycinina y ácido paraaminosalicílico
 - d) reposo en cama y aureomicina
 - e) psicoterapia y reeducación física
5. *¿Cuál será el probable desenlace de la enfermedad si no hay tratamiento?*
 - a) la fiebre desaparecerá espontáneamente con una crisis final
 - b) la convalecencia será progresiva con recaída previsible

- c) se desarrollará un empiema residual .
- d) después de la curación aparecerá una fibrosis residual
- e) puede aparecer una caverna pulmonar

g) **Comparaciones cuantitativas.**

Compare X con Y:

X — presión mecánica en el capilar venoso

Y — presión oncótica en el capilar venoso

y diga si:

A. X es mayor que Y

B. Y es mayor que X

C. X es igual a Y

h) **Relaciones.**

Sea:

1. La circulación cutánea — Y

2. La cantidad de calor perdido por unidad de tiempo

A. El aumento de la primera es acompañado de un aumento de la segunda, o bien, la disminución de la primera es acompañada de una disminución de la segunda.

B. El aumento de la primera es acompañado de una disminución de la segunda, o bien, la disminución de la primera corresponde a un aumento de la segunda.

C. Las variaciones de la segunda son independientes de las variaciones de la primera.

i) **Complementos agrupados.**

Este procedimiento se emplea cuando una pregunta puede tener más de una respuesta correcta.

Ej.: Cinco reclutas miden: 1,65 m - 1,67 m - 1,69 m - 1,63 m y 1,61 m.

1. La media de las tallas de la muestra es de 1,65 m

2. El desvío tipo es aproximadamente de 8

3. El desvío tipo es aproximadamente de 2,8

4. Es factible que la muestra informe acerca de la talla de los individuos en general

A. 1, 2 y 3 son correctos

B. 1 y 3 son correctos

C. 2 y 4 son correctos

D. 4 es correcto

E. Una sola de las proposiciones 1, 2 y 3 es correcta

En el anexo se hallará una comparación entre un examen realizado según el método tradicional y según el método de elección múltiple.

4. Cálculo de la eficacia de los "distractores"

En una pregunta de elección múltiple es importante verificar si los distractores desempeñan bien su papel. Con este propósito se calcula, para cada pregunta, qué porcentaje de las elecciones ha correspondido a cada posible respuesta.

a) Situación ideal.

A	B	C	D	E
10%	10%	10%	60%	10%

D es la respuesta correcta.

Las otras elecciones resultaron igualmente atractivas.

b) Situación por corregir.

A	B	C	D	E
3%	0%	35%	60%	2%

Solo el distractor C obró realmente. Los otros se mostraron impotentes.

5. Críticas y refutación parcial

Aunque su valor ha quedado demostrado desde hace largo tiempo, el examen con preguntas de elección múltiple sigue suscitando acaloradas críticas. Algunas son fáciles de refutar; otras parecen justificadas y no faltan las que no pueden invalidarse ni confirmarse por falta de criterios científicos.

a) Una objetividad engañosa.

La calificación de las preguntas de elección múltiple es indiscutiblemente objetiva: las respuestas correctas se definen de antemano; el alumno las ha encontrado o no. Las ventajas del método resultan evidentes.

No obstante, la subjetividad no está completamente eliminada ni mucho menos: subsiste, por lo menos en parte, en la redacción de las preguntas y en la decisión respecto de la respuesta que ha de considerarse como correcta.

En primer término, en la redacción de las preguntas. Estas son el fruto de la reflexión, de la inventiva de los examinadores, que, en última instancia, proceden en forma subjetiva. Además de las preguntas que ellos plantean, ¿cuántas otras, tal vez más válidas, hubiera sido posible imaginar?

Después, en la redacción de las respuestas que se consideran ideales. Pocas veces se trata de respuestas tan evidentemente correctas como $2 \times 2 = 4$. Durante la construcción de *tests* con nuestros estudiantes, más de una vez encontramos respuestas que se proponían como acertadas, que habían pasado victoriosamente por el tamiz del análisis matemático y que, sin embargo, no eran, en el mejor de los casos, sino burdas aproximaciones expresadas en un lenguaje dudoso.

Tal como observa Vernon, no sin malicia, la indiscutible superioridad de los exámenes "objetivos" frente a las pruebas tradicionales se debe, sin duda, con mayor frecuencia, a la preparación sumamente esmerada y a la unión de los esfuerzos y de la competencia de numerosos educadores y psicopedagogos que a la naturaleza misma del examen¹⁰.

Se olvida, además, que entre la utópica objetividad y la subjetividad total existen numerosos estadios intermedios.

b) Elecciones "correctas" controvertibles.

Este punto también fue tratado en excelente forma por Vernon, quien se expresa de la siguiente manera¹¹ :

"Es frecuente que personas altamente instruidas critiquen las preguntas de elección múltiple mediante la afirmación: 'Este ítem es estúpido', o bien que pretendan que las elecciones, consideradas como 'falsas' por la persona que elabora el *test*, resulten tan admisibles, si no más, como la respuesta 'correcta'. Esa crítica se debe principalmente al hecho de que el ítem objetivo no se basa en las mismas

¹⁰ Vernon, *The C.S.E.: An Introduction to Objective-Type Examination*, Londres, H.M.S.O., 1964, págs. 4-5.

¹¹ Vernon, *op. cit.*, pág. 6.

capacidades que las preguntas tradicionales. Sin duda, los exámenes contruidos por aficionados pueden contener no pocos ítem mal hechos. También se encuentran en las pruebas redactadas por profesionales, pero no tantas veces, porque los ítem no satisfactorios son rechazados, durante el control preliminar, por especialistas de la disciplina en consideración, o bien eliminados en oportunidad del análisis matemático de las respuestas ... Asimismo, los críticos corren el riesgo de olvidar que sus reacciones sofisticadas pueden diferir mucho de las de los alumnos con quienes fueron probados los ítem. Por último, los críticos pueden leer en las preguntas cosas que ni se les ocurrirían a alumnos inteligentes."

c) Un juego de azar.

Frente a una pregunta cerrada con dos respuestas por elegir, una correcta y la otra falsa, la probabilidad de éxito del alumno que confía en el azar es, teóricamente, de un 50%. Las personas que elaboran los *tests* saben esto desde hace mucho tiempo.

En la actualidad se trabaja, en la mayoría de los casos, con preguntas de cinco elecciones, una de las cuales es correcta y las otras solo verosímiles (distractores)¹². En ese caso, el juego de azar se reduce en forma considerable. Además, el desempeño de los alumnos variará según el nivel de sus conocimientos en la materia. En virtud de la ley de probabilidades, el estudiante que no sabe nada no tiene sino una posibilidad mínima de caer ciegamente en la solución correcta. En cambio, un conocimiento parcial permite eliminar, con conocimiento, ciertos distractores; la probabilidad de tener éxito al elegir ciegamente entre las posibilidades restantes es entonces mucho mayor que en el caso anterior, y es justo que ocurra así.

Por lo demás, existe la posibilidad de reducir en gran medida la influencia del azar mediante la corrección de los resultados obtenidos aplicando una sencilla fórmula¹³ la cual,

¹² A veces convienen varias elecciones, pero con diversos grados. Hay que seleccionar entonces la más adecuada.

¹³
$$\frac{\text{Número de respuestas correctas} - \text{Número de respuestas incorrectas}}{\text{Número de elecciones} - 1}$$

sin duda, suele castigar en forma exagerada, porque la parte correspondiente a la elección ciega varía según los alumnos.

d) Acrobacia mental.

Se reprocha a los exámenes tradicionales el hecho de que en ellos influiría demasiado la facilidad de expresión. Sin embargo, cabe preguntarse si los ítem de elección múltiple un tanto complicados (acabamos de ver algunos ejemplos) no presentan un inconveniente también grave al agregar, a la dificultad inherente a la materia de la pregunta, la obligación de desenredar negaciones dobles, de captar sutilezas lógicas o, de una manera más general, al acordar un premio a cierta aptitud de abstracción a partir de datos verbales. En todo caso, P. Vernon deja constancia de que la comprensión de la lectura desempeña un papel importante en el éxito de esas pruebas¹⁴.

Es importante conocer no solo el nivel de desarrollo mental de los alumnos, sino también saber con toda claridad lo que se quiere: verificar el conocimiento de la materia, la capacidad de comprensión en la lectura y en el razonamiento, o la combinación de los dos elementos.

No se concibe cómo podría llevarse a cabo la exploración de las cualidades intelectuales superiores sin un acrecentamiento de la complejidad de los ítem. Por esa razón, varios autores opinan que las preguntas abiertas seguirán desempeñando un papel esencial en los exámenes de nivel avanzado.

En caso de recurrir a la corrección automática, se agrega aún otra dificultad: el empleo de tarjetas de respuestas donde no suelen figurar más que los números de las preguntas y las letras A, B, C, D, E que representan las cinco elecciones. Ese sistema exige al alumno una atención adicional, sobre todo cuando no puede responder a ciertas preguntas y decide pasarlas por alto. Si se olvida de pasar por alto, asimismo, el renglón correspondiente en la tarjeta de respuestas, las consecuencias pueden ser desastrosas. Sin embargo, no hay que exagerar la dificultad: Remmers, Gage y Rummel muestra-

¹⁴ P. Vernon, "The Determinants of Reading Comprehension", en *Educational Psychological Measurement*, 1962, 22, 269-286.

ron en forma experimental que el sistema es aplicable a los niños, a partir de los 9 ó 10 años¹⁵.

De todos modos, los alumnos deben ejercitarse cuidadosamente en la técnica de los exámenes de elección múltiple antes de someterse a una prueba decisiva. Con esta precaución y mediante una construcción rigurosa, las preguntas de elección múltiple proporcionan buenos resultados; su utilización intensiva en los países anglosajones durante muchos años y su difusión cada vez mayor en otras partes constituyen una prueba elocuente.

e) Inconvenientes inciertos.

Calificamos de inciertos los siguientes reproches porque, según sabemos, ninguna investigación científica rigurosa estableció sus fundamentos. Aparentemente, ciertas críticas solo se justifican en la medida en que el examen está mal construido.

- Elegir la respuesta correcta es más fácil que elaborarla. Interviene demasiado la memoria. Se corre el riesgo de que se instalen cierta pereza intelectual, una repugnancia al esfuerzo y a la formulación clara del pensamiento.
- En conocimiento de que, para rendir el examen de lenguaje, bastaría con subrayar la respuesta elegida o marcarla con una cruz, los maestros harían caso omiso de los ejercicios de expresión.
- Se daría preferencia a numerosos pequeños ejercicios realizables en muy poco tiempo, y se dejarían de lado, de esa manera, los problemas que requieren una larguísima reflexión. Ahora bien, los estudios superiores y la vida real no nos ahorran ese tipo de pruebas.

C. A manera de conclusión: un compromiso

Nos parece que se impone un compromiso, en dos niveles por lo menos.

Como tanto las preguntas abiertas cuanto las cerradas parecen poseer sus ventajas particulares y poner a prueba

¹⁵ H. Remmers, N. Gage y Rummel, *Educational Measurement and Evaluation*, Nueva York, Harper, 1955, pág. 246.

diferentes aprendizajes, no vemos razón alguna para reclamar un monopolio con respecto a alguno de los dos tipos. Pueden coexistir sin problemas. Sin embargo, los mejores docimólogos opinan que no conviene mezclarlos en una misma prueba.

Por lo demás, existe una fórmula intermedia entre preguntas abiertas y cerradas que tal vez se imponga cada vez más, según los adelantos de la investigación pedagógica.

El examen consistiría en un número bastante grande de preguntas *abiertas*, previamente "testeadas", y de un alcance relativamente limitado. Se conocería, pues, de antemano, el abanico de las respuestas probables; esto permitiría proponer un esquema de calificación que se acercaría mucho al rigor de la pregunta con respuesta cerrada.

Esta posibilidad ya fue confirmada por varios trabajos experimentales. De esta manera se obtuvieron correlaciones casi perfectas (0,98 y 0,99) entre siete grupos de gente encargada de calificar que representaban, cada uno, una comisión de exámenes de Gran Bretaña. La experiencia se hizo con pruebas de historia de nivel superior al G.G.E. (fin del ciclo secundario)¹⁶.

Sin ir tan lejos, los profesores pueden ponerse de acuerdo perfectamente acerca de los puntos por exigir en las respuestas a preguntas abiertas.

Por lo demás, la manera de redactar las preguntas puede cerrar de modo parcial las respuestas y disminuir, en consecuencia, el juego de la subjetividad. Una pregunta como "Escriba un comienzo adecuado a la frase (. . . cuando empezó a llover)" se halla, de alguna manera, a mitad de camino entre la prueba subjetiva y la objetiva, pues aquí serán pocos los casos de desacuerdo entre los encargados de calificar que deban pronunciarse con respecto a la corrección de la respuesta.

¹⁶ JOINT MATRICULATION BOARD, *The Marking of Scripts in Advanced Level History*, Universidades de Manchester, Liverpool, Leeds, Sheffield y Birmingham, 1964.

LA CALIFICACION

I. Un preámbulo indispensable: la curva de Gauss

Es imposible hablar de evaluación de los resultados sin recurrir a algunas nociones matemáticas. No olvidamos la promesa hecha al comienzo de esta obra: ¡la aritmética elemental será suficiente!

A. La curva de Gauss, imagen de la probabilidad

En una bolsa hay dos bolitas completamente iguales, solo que una es roja y la otra blanca.

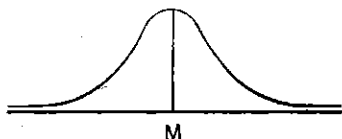
En esas condiciones, cada bolita tiene una probabilidad entre dos de salir en cada tiro hecho a ciegas.

Si nos entregamos a ese juego con un número limitado de tiros, es posible que saquemos varias veces consecutivas el mismo color. Esa probabilidad disminuye a medida que aumentamos el número de tiros.

Imaginemos 100 tiros consecutivos. Es muy probable que el rojo salga aproximadamente tantas veces como el blanco. En cambio, es muy poco probable que obtengamos 100 veces consecutivas la bola roja o la blanca. Ya es un poquito más probable obtener 99 veces rojo y una vez blanco; más probable aún 98 veces rojo y 2 veces blanco, etcétera. En síntesis, la probabilidad aumentará hasta 50 veces rojo y 50 blanco, para disminuir luego de manera progresiva, hasta llegar a 1 vez rojo y 99 veces blanco.

En un número muy grande de tiros, ese movimiento ascendente-descendente corresponde a una curva en forma de campana: es la famosa curva de Gauss, cuyas dos mitades son

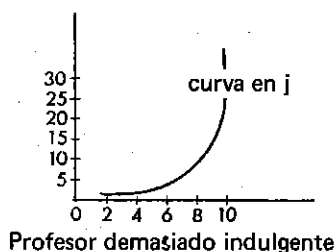
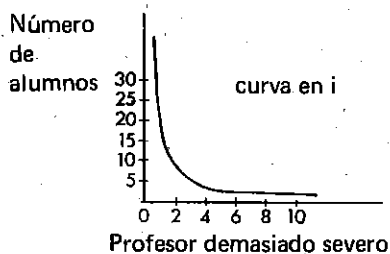
simétricas en relación con la media aritmética y cuyos extremos no tocan nunca la línea del cero, pues la probabilidad nula solo existe en el infinito.



Esa *distribución*, llamada *normal*, refleja la imagen de muchas cualidades humanas, tal como se distribuyen en grupos numerosos, *tomados al azar*. De este modo, entre los habitantes de una gran ciudad, los hombres de talla mediana son los más numerosos, mientras que los gigantes y los enanos resultan muy raros; entre estos dos extremos, la población se distribuye según la curva de Gauss.

Hasta el error se somete muchas veces a la ley normal: si hacemos una cantidad muy grande de mediciones, es muy probable que veamos aparecer un error de magnitud media y una distribución que va del error ínfimo al máximo. Veremos más adelante que esta observación es importante.

Cuando deja de intervenir el azar, la distribución se modifica. Por ejemplo, un profesor puede poner muchísimas notas malas y muy pocas buenas, o al revés. Es factible que se obtenga entonces una de estas dos curvas características:



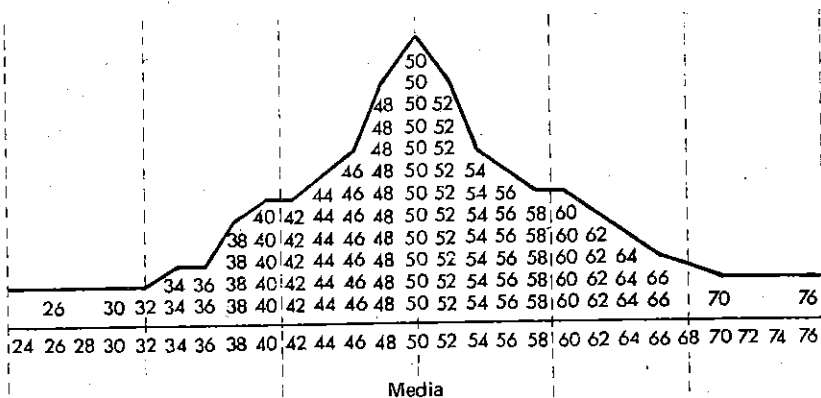
B. La curva de Gauss, imagen de los resultados de la enseñanza no individualizada

Un profesor que enseña de manera no individualizada en una clase en la que los alumnos no están especialmente seleccionados, suele dar a su curso un grado de dificultad adaptado a la mayoría del grupo. Si el ajuste es correcto, obtendrá muchos resultados medianos, pocos muy buenos y escasos muy malos. La distribución de esos resultados se acercará a la curva de Gauss.

Un fenómeno similar se produce, en forma todavía más espontánea, en los ejercicios que se sustraen a la cuantificación rigurosa, porque ponen en juego un complejo conjunto de factores.

A continuación dispondremos 100 resultados (imaginarios) obtenidos en un examen calificado con 100 puntos. Vemos que 12 alumnos obtuvieron 50 de 100 (nota promedio), mientras que uno solo tiene la nota más baja (26) y uno la más alta (76).

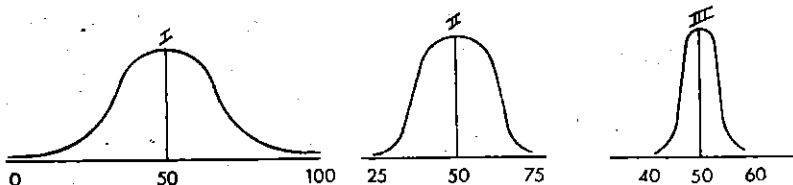
Observamos, además, que la línea que corresponde a la distribución de los resultados no es una curva de campana perfecta, sino tan solo en forma aproximada. En tales casos (pronto ofreceremos algunas precisiones al respecto) se supone que si el número de sujetos hubiese sido más elevado la curva se habría alisado; la distribución se considera "normal".



C. El desvío tipo (estándar) o sigma, índice precioso

1. Significado

La variación de los 100 resultados que acabamos de examinar hubiera podido ser más grande, por ejemplo de 0 a 100, o más pequeña, por ejemplo de 40 a 60, siempre distribuyéndose según la curva de Gauss. Hubiéramos podido tener:



En los tres casos, la media es 50. No obstante, se trata de situaciones muy diversas.

El margen de variación de los resultados es:

$$\text{curva I : } 100 - 0 = 100$$

$$\text{curva II : } 75 - 25 = 50$$

$$\text{curva III: } 60 - 40 = 20$$

Sigma o σ es un índice fácil de calcular (veremos más adelante el método de calcularlo) que nos indica de inmediato la forma en que varían los resultados en relación con la media aritmética de las notas obtenidas.

Para las 100 notas tomadas como ejemplo, sigma = 9. Si consideramos un sigma antes y después de la media, obtenemos: 41 y 59. Mediante el recuento del número de resultados que caen entre esos dos límites obtenemos 68 notas, o sea aquí un 68% de las notas de todo el grupo.

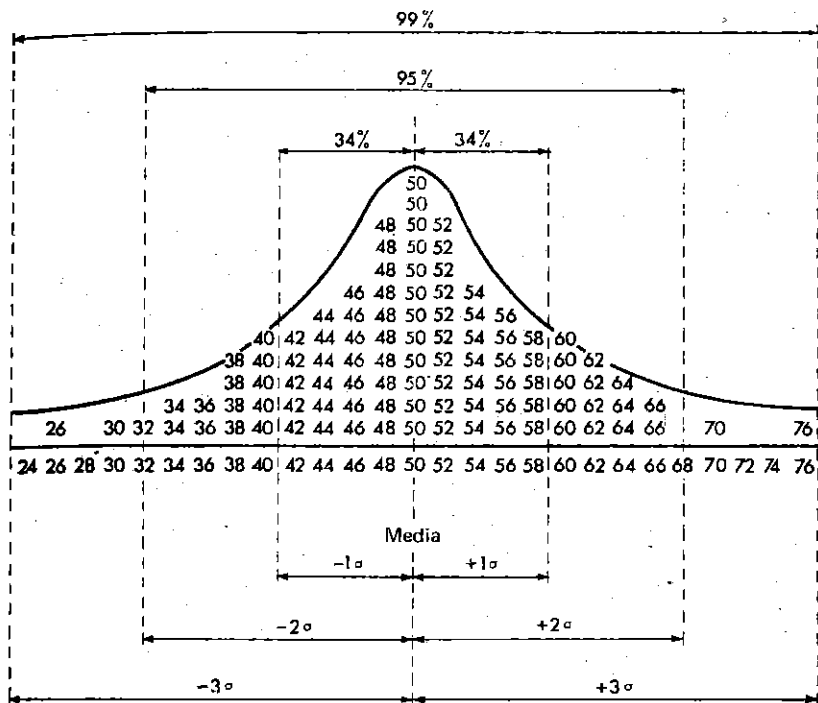
Siempre ocurre de esta manera en una *distribución normal*: sigma indica invariablemente la misma proporción de los resultados en relación con el conjunto:

1 sigma a cada lado de la media = 68% de las notas

2 sigma a cada lado de la media = 95% de las notas

3 sigma a cada lado de la media = 99% de las notas

Veamos ahora con nuevos ojos la distribución de los resultados:



Como a un determinado sigma corresponde siempre la misma superficie del área debajo de la curva, es decir, la misma proporción de los resultados, *ese índice nos suministra el medio de comparar los resultados atribuidos a distintos profesores, en cuanto esos resultados se distribuyen en forma normal*. Pronto mostraremos cómo.

2. Estimación rápida de la media y del sigma

a) Problema.

El cálculo exacto de la media y del desvío tipo (sigma) resulta relativamente difícil. Además, si la distribución no es

normal, la media aritmética ofrece una idea equivocada de la tendencia central.

Un procedimiento sencillo permite evitar largos cálculos. Solo proporciona resultados *aproximados*. Sin embargo, éstos son suficientes en casi todos los casos de la práctica cotidiana en la escuela¹.

Tomemos como ejemplo dos preguntas de examen a las cuales respondieron 10 alumnos. He aquí las notas atribuidas, las medias y los desvíos tipos calculados según el método clásico:

<i>Alumno</i>	<i>Pregunta 1</i>	<i>Pregunta 2</i>
1	39	32
2	33	28
3	25	32
4	22	28
5	26	27
6	18	31
7	23	33
8	13	27
9	57	35
10	45	36
	301	309

Media
aritmética: 30
Desvío tipo:

$$\sigma = \sqrt{\frac{\sum d^2}{N}} = \pm 13.$$

Media
aritmética: 31

Desvío tipo:
 $\sigma = \pm 3.$

¹ Según D. R. Mather y colab., *op. cit.*

b) *Método sencillo de cálculo.*

1° Representación gráfica de los resultados.

Notas	Pregunta 1	Pregunta 2
60		
58		
56		
54		
52		
50		
48		
46		
44		
42		
40		
38		
36		
34		
32		
30		
28		∴∴
26	:	
24		
22	:	
20		
18		
16		
14		
12		
10		

Márgenes de variación:

Pregunta 1: $57-13 = 44$

Pregunta 2: $36-27 = 9$

Las notas de la pregunta 1 están mucho más dispersas. Influyen más en la clasificación final que las de la pregunta 2. A menudo conviene pasar por alto las dos notas extremas; en este caso:

Margen pregunta 1 = 27
pregunta 2 = 8

Medianas

= notas del medio.

Aquí, número par; por lo tanto la media entre 5° y 6°.

Pregunta 1:

$(25 + 26) : 2 = 25,5$

Pregunta 2:

$(31 + 32) : 2 = 31,5$

En cuanto a la pregunta 1, la diferencia bastante nítida entre la media aritmética y la mediana (30 y 25,5) indica una distribución asimétrica de las notas (aquí: la mayoría debajo de la mediana).

En cuanto a la pregunta 2, la simetría es buena (media 31 y mediana 31,5).

Observemos que en una distribución perfectamente normal la media y la mediana coinciden: véase nuestro ejemplo.

2° Estimación del desvío (σ).

$\sigma = \frac{1}{2}$ de la diferencia intercuartil.

Por lo tanto

- a. Buscar el cuartil superior, es decir, el medio de las notas por encima de la mediana.

$$\text{Pregunta 1} = 39$$

$$\text{Pregunta 2} = 33$$

- b. Buscar el cuartil inferior

$$\text{Pregunta 1} = 22$$

$$\text{Pregunta 2} = 28$$

- c. Diferencia intercuartil.

$$\text{Pregunta 1} = 39 - 22 = 17$$

$$\text{Pregunta 2} = 33 - 28 = 5$$

- d. Estimación del σ :

$$\text{Pregunta 1} = 17 \times \frac{1}{2} = 12\%$$

$$\text{Pregunta 2} = 5 \times \frac{1}{2} = 3\%$$

D. La concentración de los resultados alrededor de la media

Volvamos una vez más a la distribución de las 100 notas tomadas como ejemplo (pág. 99). Observamos que, cuanto más nos acercamos a la media tanto más numerosas son las notas:

12 notas de 100 en la media exacta

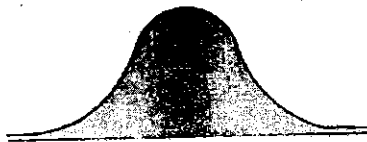
22 notas de 100 entre 48 y 52

68 notas de 100 entre 41 (-1σ) y 59 ($+1 \sigma$)

Imaginemos que se trata de redacciones en francés. Mientras el profesor podía calificar de 0 a 100, casi una cuarta parte de las notas se sitúan entre 48 y 52, es decir, a 4 puntos de distancia.

¿Qué significa una diferencia de 4 puntos sobre 100 en una redacción francesa? Es casi seguro que, si hacemos juzgar el mismo trabajo por diez profesores diferentes, los desniveles serán casi siempre superiores a 2. Ahora bien, si la media constituye el límite del fracaso (aquí coincide con la mitad de la escala total), puede afirmarse, sin temor a equivocación, que para un 22% del grupo de alumnos tomados como ejemplo, el fracaso o el éxito dependerán simplemente del azar.

Aquí tenemos, en degradación, una imagen de la densidad de la población en relación con la media:



E. Curva de Gauss reclamada por los maestros

La curva de Gauss constituye un excelente instrumento de calificación porque permite identificar a los mejores (por ejemplo, entre 25 y 35) y a los peores.

Cuando se trata de un concurso en que numerosos candidatos se disputan unos pocos puestos, es factible, por ejemplo, elegir entre los que se ubican entre $+1\sigma$ y $+3\sigma$ (la parte que se halla fuera de la fuerte concentración alrededor de la media). Así también, si se desea calificar a los alumnos —lo que se hacía demasiado exclusivamente en el pasado, pero que sigue siendo útil, a pesar de todo, para permitir a cada uno ubicarse— la distribución *normal* es, sin duda, provechosa.

Como esta vez ya no se trata de grupos elegidos al azar, sino de individuos ejercitados en forma especial, el profesor *crea artificialmente* las condiciones necesarias para una distribución según la curva de Gauss. ¿Cómo? Dosificando los ítem de acuerdo con sus índices de dificultad (y de eficacia).

F. ¿Cómo se sabe si una distribución es normal?

Todo lo que acabamos de expresar a propósito de la curva de Gauss y del sigma sólo es válido si los resultados se distribuyen en forma *normal*. Antes de entrar en alguna operación es importante, en consecuencia, verificar si esa condición se cumple.

Existen rigurosos procedimientos matemáticos para hacerlo, pero, una vez más, nos conformaremos con una simple aproximación gráfica: el *histograma*.

El procedimiento es sencillo:

1. *Clasificar los resultados.*

Existe la posibilidad de ordenarlos de mayor a menor, o viceversa.

He aquí lo que obtenemos al clasificar los 100 resultados que ya tomamos como ejemplo:

<i>Notas</i>	<i>Número de alumnos que obtuvieron estas notas (frecuencias) (f)</i>	
26	I	1
30	I	1
32	I	1
34	II	2
36	II	2
38	IIII	4
40	IIII	5
42	IIII	5
44	IIII I	6
46	IIII II	7
48	IIII IIII	10
50	IIII IIII II	12
52	IIII IIII	10
54	IIII II	7
56	IIII I	6
58	IIII	5
60	IIII	5
62	IIII	4
64	III	3
66	II	2
70	I	1
76	I	1

N = 100

Pero si las notas son numerosas, este procedimiento no es rápido ni práctico. Es mucho más fácil constituir clases.

1. Calcular el margen de variación entre las dos notas extremas: $76 - 26 = 50$.
2. Dividir la diferencia por 15^2 : $50:15 = 3,33$.

² Ese número es arbitrario. La práctica muestra que, en la mayoría de los casos, lleva a una correcta distribución.

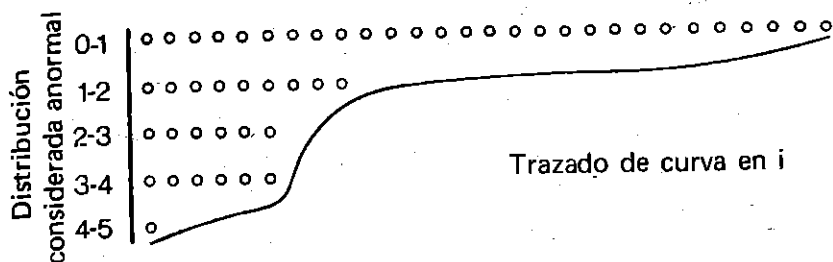
3. Elegir como intervalo de clase uno de los dos números impares más cercanos: 3 ó 5. En vista de la poca cantidad de notas, elegimos aquí 5.
4. Colocar la nota más alta en el medio del intervalo superior. Nota más alta: 76; la clase superior es, pues: 74-75-76-77-78.

	Clases	Frecuencias (f)
1	74 - 78	1
2	69 - 73	1
3	64 - 68	5
4	59 - 63	9
5	54 - 53	18
6	49 - 53	22
7	44 - 48	23
8	39 - 43	10
9	34 - 38	8
10	29 - 33	2
11	24 - 28	1
11 clases	Intervalo (i) = 5	Número de notas (N) = 100

2. Dibujar el histograma.

En la parte "frecuencia" del cuadro anterior, el puntaje se hizo con mucho cuidado. Basta con tener la página en sentido horizontal para que aparezca el esqueleto del gráfico denominado histograma: solo quedan por trazar los rectángulos.

Si unimos los centros de los extremos superiores, obtenemos una línea (polígono de frecuencias) que, en este caso, es un esbozo con el cual nos conformaremos para considerar que la distribución es, muy probablemente, normal. ¿Por qué? Porque el trazado tiene *un solo vértice*, casi *en el centro*, a partir del cual las cantidades *van disminuyendo con bastante regularidad por ambos lados*.



II. La calificación subjetiva: escala de evaluación

A. Introducción

Por calificación subjetiva comprendemos toda evaluación realizada por los maestros de una manera global, ya sea de manera inmediata a la observación del rendimiento de los alumnos, o bien después de haber obtenido una visión de conjunto de puntajes particulares. De modo más general, puede afirmarse que es subjetiva toda nota a la cual se llega por un procedimiento no susceptible de automatización.

El grado de subjetividad varía considerablemente. A veces apenas si tiene límites (por ejemplo: "Esto está bien... —Esto está muy bien..."); otras veces casi cede el lugar a una técnica objetiva (ejemplo: una escala cuyos grados se definen con una precisión que prácticamente no deja ya lugar a la interpretación personal).

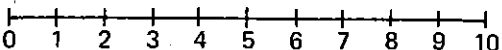
En todos los casos, la evaluación subjetiva se realiza por comparación con criterios externos o internos. La persona que califica llega a la conclusión de que hay coincidencia, o no, en relación con el criterio. Establece, pues, una clasificación.

Por esa razón, podemos limitar esta parte al estudio de las escalas de evaluación.

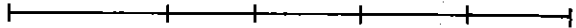
La escala de evaluación ayuda a *ordenar* los individuos o los objetivos en relación con una cualidad que poseen en un grado más o menos elevado.

Representa un *continuo* que, teóricamente, puede extenderse desde un mínimo absoluto hasta un máximo absoluto, pero que en la práctica no suele ser más que una fracción, un segmento de ese absoluto. De lo excelente... a... lo insuficiente, de lo muy pequeño... a lo muy grande, etcétera.

La escala de evaluación, gráfica o no, es:

numérica: 

o bien *descriptiva*: muy bien, bien, regular, mediocre, insuficiente



La escala de evaluación es el instrumento más frecuentemente usado en la escuela... y el menos comprendido. Podemos afirmar, sin temor a equivocarnos, que ha viciado literalmente gran parte de los exámenes en todos los niveles escolares y, probablemente, ha perjudicado a un número considerable de personas.

Por eso resulta absolutamente esencial conocer su verdadera naturaleza y, sobre todo, sus enormes limitaciones.

B. Indole y debilidades de las escalas de evaluación

La escala de evaluación es *una escala ordinal*.

- 1) *No posee ni cero ni máximo natural o absoluto; empieza y termina en cualquier lugar elegido por quien la emplea.* Por consiguiente, aunque evalúen un mismo fenómeno desde un idéntico punto de vista, los jueces nunca tienen escalas rigurosamente iguales; las diferencias pueden ser considerables y muchas veces lo son.

Ejemplo: *Conocimiento del inglés como idioma extranjero.*

Teóricamente, el cero podría ser el momento en que el alumno aún no conoce una sola palabra del idioma. Sin embargo, pocas veces se juzga en la escuela a partir de ese punto.

En cuanto al conocimiento total, absoluto, es evidente que no existe.

Imaginemos dos profesores de inglés que juzgan, sin ponerse de acuerdo, dos divisiones en las que cada alumno ha hecho una breve exposición oral. Lo más probable es que los dos extremos de las escalas de evaluación, utilizadas independientemente por los dos profesores, resulten definidos en mayor medida por el alumno juzgado como el más capaz y el considerado como el más flojo de

su grupo, que por una evaluación abstracta de la cantidad de inglés que deberían haber asimilado en el nivel pedagógico considerado.

- 2) *Los grados no son de igual magnitud dentro de una misma escala.*

Por ejemplo, es imposible demostrar que la distancia que separa el conocimiento bueno del muy bueno, en inglés, es la misma que separa el conocimiento regular del bueno.

Si los grados son A, B, C, D, E, no es posible afirmar, pues, que $A - B = B - C = 0$

- 3) *Los grados correspondientes de una escala a otra (para una misma materia) no son de igual magnitud.*

Bueno en inglés, para un profesor, no es cuantitativamente lo mismo que bueno en inglés para otro: nadie puede establecer semejante igualdad en forma matemática.

- 4) *Escalas referidas a diferentes materias no son comparables.*

Para los alumnos de un mismo grado de la escuela primaria, donde un solo maestro enseña todas las materias, ser *bueno* en matemáticas no abarca los mismos fenómenos que ser *bueno* en francés. En un plano cuantitativo, los escalones no son, por cierto, comparables.

De todo esto se deduce que las escalas de evaluación no permiten ninguna operación aritmética. Si se califican la redacción francesa y el inglés como idioma extranjero de 0 a 10, nada permite sumar los dos resultados para evaluar un saber total. No obstante, esto es lo que se hace, desde hace decenios, en nuestras escuelas. Es igualmente sospechoso, si no más, utilizar escalas numéricas diferentes para las diversas materias, en función de una ponderación empírica: de 0 a 120 para el francés, de 0 a 80 para las ciencias, etcétera.

Si el sistema ha funcionado, esto se debe a que ha permitido, de todos modos, identificar a los mejores (aquellos que se clasifican *en todo* entre los primeros) y los más flojos (los últimos *en todo*).

C. Utilidad

A pesar de esas limitaciones, la escala de evaluación es, en muchos casos, el único instrumento de que disponemos para concretar nuestro juicio acerca de complejos comportamientos humanos.

De hecho, permite aportar a un legajo escolar elementos *cualitativos*, tan objetivos como sea posible, *complementos indispensables* de los resultados *cuantitativos* aportados por los exámenes objetivos o los *tests*.

D. Construcción

Muchos educadores se dejan engañar por la facilidad aparente con que se construye una escala.

En realidad, no es raro que la preparación de tal instrumento requiera largos meses de trabajo, imprescindibles para la aclaración teórica, y que exija el empleo de técnicas refinadas como el análisis factorial. Este ayuda, sobre todo, a determinar en qué medida varias escalas evalúan las mismas cosas con diversas denominaciones.

También en este dominio es importante, en consecuencia, que los docentes puedan recurrir cuanto antes a servicios de investigación bien equipados de especialistas y computadoras.

Sin embargo, esto no significa que los maestros no puedan hacer algo mientras aguardan.

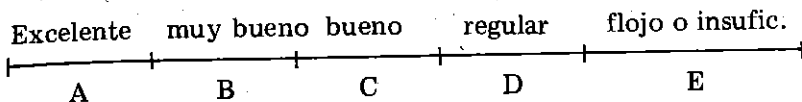
1. ¿Cuántos grados?

Teóricamente, una escala de evaluación puede contar con una cantidad infinita de escalones. En la práctica, los jueces solo califican con alguna seguridad si se emplea un número reducido de grados: 3, 5, 7 ó 9.

Refiriéndonos a la literatura experimental y a múltiples observaciones que pudimos hacer en forma directa, aconsejamos atenerse *en la mayoría de los casos a 5 escalones*. Estos permiten calificaciones relativamente seguras y fieles, siempre que cada escalón esté muy claramente definido y los encargados de la calificación hayan sido, por consiguiente, bien preparados.

Desde el punto de vista práctico:

1. Se califica primeramente en tres categorías: los mejores, los más flojos, el resto.
2. Después se distribuye "el resto" entre buenos, regulares y flojos



Observemos que el escalón C abarca el centro de la escala, es decir, la media aritmética. Tal como hemos visto en oportunidad de la curva de Gauss, muchas veces se produce una fuerte concentración alrededor de ese punto. Elegir un número par de escalones crearía un centro en la escala, delante o detrás del cual tambalearían varios alumnos abandonados al azar.

En caso de una distribución normal, no queda excluido el hecho de que el veredicto del azar afecta a un 20, 30 e incluso a un 40% de los alumnos.

2. Definir el objeto de evaluación

Pedir que se aprecie, por ejemplo, el "coraje" de los miembros de un determinado grupo —sin ninguna otra precisión— conduciría a respuestas casi totalmente carentes de interés. ¿Qué significado, o cuáles significados, habrán otorgado los observadores a la palabra "coraje": ardor, voluntad, celo, perseverancia, valentía, firmeza, estoicismo?

Y aunque precisáramos que por "coraje" entendemos la firmeza frente al peligro, ¿nos sería fácil distinguir entre intrepidez y temeridad?

Con el fin de mitigar las imprecisiones de calificaciones tales como *muy bueno*, *regular*, etc., es necesario agregar a la escala una descripción tan precisa como sea posible acerca del rasgo u objeto por apreciar, y ejemplificar la definición con situaciones tipo. Definir, no mediante formulas abstractas, sino por medio de comportamientos precisos, es condición *sine qua non* de validez.

He aquí cómo Schonell³ presenta el rasgo "confianza en sí mismo":

Confianza en sí mismo

Extremadamente confiado en sí mismo. Casi demasiado seguro de sí.	Muy confiado en sus propias fuerzas.	Confiado.	Falta de confianza. Tímido.	Falta extrema de confianza. Depende de los otros. Evita las responsabilidades.
---	--------------------------------------	-----------	-----------------------------	--

Descripción del rasgo

"En su forma positiva, esta cualidad se caracteriza por las siguientes manifestaciones: el individuo cuenta consigo mismo, es capaz de enfrentar las dificultades, tiene seguridad, es independiente y está dispuesto a asumir responsabilidades.

El niño que tiene confianza en sí mismo trata de avanzar con un mínimo de ayuda; el que carece de confianza requiere ayuda constantemente. Al primero le gusta ver que es capaz de construir y producir cuando recibió instrucciones claras; el segundo quiere que se le prepare la tarea y que se le ayude durante toda la etapa de realización.

Situaciones tipo

1. ¿Tiene miedo de la oscuridad?
2. ¿Sabe cuidar de sí mismo y de sus cosas o es necesario que siempre tenga alguien a su lado?
3. ¿Viaja solo en tranvía o en ómnibus? (Niños mayores de 9 años.)
4. ¿Habla con desenvoltura a visitas desconocidas?
5. ¿Es bueno en los juegos? ¿Sabe nadar?
6. ¿Se siente cómodo y responde con seguridad en los exámenes orales?
7. ¿Lee bien, dramatiza correctamente un texto frente a la clase?

³ F. J. Schonell, *Backwardness in the Backward Subjects*, citado por F. Warburne, *Measurement of Personality* ("Educational Research", nov. 1961, pág. 9).

8. ¿Aborda bien las tareas nuevas o hace continuamente preguntas a sus compañeros y maestros?

Definir los grados de la escala

Ejemplo: Organización de la enseñanza de la lectura.

0	1	2	3	4	5
Mediocre En lectura todos los alumnos siguen la misma progresión. No hay trabajo en grupos.	Bastante buena Cf. 1. Pero a veces un alumno muy lento recibe un poco menos de trabajo que los otros.	Regular Se constituyen 2 ó 3 grupos según las aptitudes en la lectura. Poca flexibilidad en el agrupamiento.	Muy buena Agrupamiento según aptitudes. Flexibilidad.	Excelente Agrupamiento después de un estudio profundo de las aptitudes y dificultades encontradas. Gran flexibilidad.	

El lector habrá notado que este último ejemplo combina las escalas gráfica, numérica y descriptiva.

El peligro de una definición insuficiente de los cinco grados fue demostrado experimentalmente por R. Weiss⁴. Al adoptar las escalas descriptivas impuestas por el Ministerio de Educación nacional de Austria, Weiss invitó a 92 maestros a calificar dos redacciones de alumnos del nivel en que ellos enseñaban (4º grado primario) y a 272 maestros a calificar dos trabajos de aritmética (problemas; 4º y 5º año).

Los siguientes diagramas muestran cuánto divergen las calificaciones para un mismo trabajo (dispersión de las notas a lo largo de toda la extensión de la escala).

CALIFICACION DE DOS REDACCIONES

	25%	50%	75%		Nota media	Sigma	
Ortografía	1	2	3	4	5	2.89	1.04
Estilo	1	2	3	4	5	2.29	0.53
Fondo	1	2	3	4		2.08	0.79
Nota global	1	2	3	4		2.45	0.80

⁴ R. Weiss, *op. cit.*

II

						Nota media	Sigma
		25%	50%	75%			
Ortografía	1	2	3	4	5	2.99	0.96
Estilo	1	2		3	4	2.12	1.05
Fondo	1	2		3	4	1.78	0.85
Nota global	1	2	3	4		2.54	0.80
		25%	50%	75%			

Calificación de dos trabajos de aritmética

						Nota media	Sigma
		25%	50%	75%			
I	1	2	3	4		2.56	0.77
II	2	3	4		5	3.55	0.80

Según vemos, no basta con adoptar una nueva escala, en sustitución del sistema tradicional de notas, para que todos los problemas queden resueltos. ¡Aún falta mucho para ello!

E. Utilización

1) ¿Cuántos alumnos por escalón?

Primer caso: El alumno comparado consigo mismo.

Los escalones marcan las etapas por superar en el ascenso hacia un poder o saber. El maestro ha definido un objetivo por alcanzar. En una enseñanza no individualizada, ese objetivo es común a la mayoría, si no a la totalidad de los alumnos de la división. Es evidente que debe elegirse de tal manera que *todos* puedan lograrlo. Será, pues, un objetivo amplio, y esto puede redundar en menoscabo de su precisión, hasta el punto de volverlo inoperante.

Está claro que actualmente procedemos como si al comienzo del año escolar todos los alumnos se encontraran aproximadamente en un mismo punto de la escala del saber. Suponemos, en consecuencia, que en el año anterior todos han alcanzado el grado más alto de la escala precedente.

Ahora bien, advertimos claramente que la realidad es otra, y la investigación confirma nuestro sentimiento:

a) *Márgenes de variación de la edad mental*

Al revisar el *test* de inteligencia de Binet-Simon en Estados Unidos, Terman y Merrill (1937) comprobaron⁵:

- que en el primer grado de la escuela primaria la edad mental variaba de 4 a 8 años;
- que en sexto grado primario la edad mental oscilaba entre 8 y 16 años;
- que en el colegio secundario, un margen de variación de 8 a 10 años no era excepcional.

b) *Márgenes de variación del rendimiento escolar*

Basta con observar cuánto se superponen las normas de los *tests* de conocimientos, graduadas por año escolar, para darse cuenta del alcance de los márgenes de variación.

Basado en sus propias investigaciones, así como en las de Lindquist, Cornell, Learned y Wood, W. W. Cook⁶ proporciona las siguientes indicaciones:

- Para la comprensión de la lectura, el vocabulario, las ciencias, la geografía y la historia, el margen de variación del rendimiento es:
 - En 1° grado de primaria: de 3 a 4 años.
 - En 4° grado de primaria: de 5 a 6 años.
 - En 6° grado de primaria: de 7 a 8 años.
- Para el razonamiento aritmético y el cálculo:
 - En 6° grado de primaria: de 6 a 7 años.
- En cultura general (ciencias, literatura, artes, historia, geografía y lenguaje) a nivel universitario:
 - 28% de los estudiantes de 4° año están por debajo de la media de los estudiantes de 2° año;
 - 10% de los estudiantes de 4° año están por debajo de la media de los alumnos que egresan del colegio secundario.

⁵ Cf. Q. McNemar, *The Revision of the Stanford-Binet Scale*, Boston, Houghton-M., 1942.

⁶ W. W. Cook, "The Functions of Measurement in the Facilitation of Learning", en E. F. Lindquist, Ed., *Educational Measurement*, Washington, A. C. E., 1961, 4a. ed., págs. 3-47.

Tales observaciones —que no han de tomarse al pie de la letra, pero cuya tendencia hay que tener presente— podrían multiplicarse a gusto.

Sirven para probar que mientras nuestras escuelas funcionen de acuerdo con el sistema de divisiones rígidas, la evaluación continua, que corre paralela a la preocupación de permitir a cada alumno que progrese según su propio ritmo, se supere siempre y llegue tan lejos como pueda, resultará una quimera.

Sin embargo, como por razones tecnológicas y económicas la enseñanza individualizada no se generalizará antes de que trascurren muchos años, solo la enseñanza parcialmente individualizada ofrece una solución realista de aplicación inmediata. En una escuela a medias individualizada, grupos homogéneos, según las principales materias, remplazan a la división. De esta manera, un alumno bien dotado para idiomas extranjeros puede trabajar en esa asignatura en un nivel A, pero pasar al nivel D para los problemas de aritmética⁷.

Con el propósito de no sobrecargar este texto, ofrecemos en el anexo la descripción detallada de una escuela que desde hace cincuenta años funciona de acuerdo con ese principio.

Segundo caso: Alumnos comparados entre sí.

Aun cuando tratemos de llevar a cada uno a un nivel elevado, probablemente persistirán bastantes diferencias para que aparezca una nueva distribución según la curva de Gauss (o por lo menos su bosquejo), donde la media simplemente se habrá desplazado hacia arriba.

Parece justificado, en consecuencia, tratar de poblar cada escalón de acuerdo con porcentajes que se aproximen a la distribución normal.

En una división habitual, evidentemente no es posible respetar en forma estricta esa proporción, a causa del reducido número de alumnos. Tendremos, por ejemplo:

⁷ No es casual que hablemos de *problemas de aritmética* y no de aritmética en general. En una materia compleja como ésta, un alumno puede estar bien dotado para el cálculo, pero ser flojo en cuanto a problemas. La investigación actual muestra que si las agrupaciones no son suficientemente detalladas, los márgenes de variación entre alumnos siguen siendo considerables.

La distribución
normal sería:

Excelentes:	1	5%	2,5%
Muy buenos:	4	20%	13,5%
Buenos (regulares):	10	50%	68 %
Flojos:	4	20%	13,5%
Muy flojos:	1	5%	2,5%
	<hr/>	<hr/>	<hr/>
	20	100%	100%

Insistimos en que obligarse a respetar semejante distribución (*distribución forzada*) no implica necesariamente el fracaso para cierto número de alumnos. Todas las notas son *relativas* unas con respecto a otras: "muy flojo" puede, en algunos grupos fuertes, ubicarse por encima de la nota que significa fracaso.

Por esa razón, *resulta más conveniente adoptar letras de calificación* en lugar de las fórmulas tradicionales.

2) Luchar contra la contaminación y la tendencia central

Además de los peligros de estereotipia y del efecto de halo, descritos en la página 27, otro fenómeno de contaminación, de índole más mecánica, acecha a los jueces. Si utilizan de manera consecutiva, para un mismo alumno, una serie de escalas orientadas en el mismo sentido (por ejemplo, de lo positivo a lo negativo o del mejor al más flojo), y tal vez presentadas incluso en una misma página, las calificaciones tienden a agruparse del mismo lado.

Se aconseja que se haga figurar cada escala en distinta página y se sortee el sentido de presentación de cada una.

Por último, conviene que pase un intervalo bastante largo después de la evaluación de cada cualidad de un mismo sujeto, y que se haga evaluar la misma calidad por el mayor número de jueces que sea posible.

Muchos jueces tienden, asimismo, a agrupar sus notas hacia el centro de la escala. Temen sobrestimar o subestimar a un alumno y afrontar la responsabilidad de un fracaso. En ese sentido, la regla de la distribución forzada presta buenos servicios. Pero para que sea realmente eficaz ha de apoyarse

en una descripción precisa y matizada de los distintos escalones.

En el caso de que las evaluaciones de varios jueces deban combinarse, los profesores que se dejan seducir por la solución fácil, que es la nota central, han de saber que renuncian a gran parte de su influencia, a favor de sus colegas que se atreven a utilizar toda la escala.

F. ¿Cómo sintetizar las evaluaciones?

Al término de un período escolar es importante sintetizar las evaluaciones (que ojalá resulten tan numerosas como sea posible), pues en caso contrario ni los maestros ni los alumnos sabrán interpretarlas.

Volver simplemente a calcular la media podría desvirtuar los esfuerzos de rigor anteriores. Porque las evaluaciones ordinales, por lo menos teóricamente, no admiten suma ni resta.

Aquí tenemos un método sencillo que permite ubicar a un alumno en cada materia y, después, en el conjunto de las materias.

Al comienzo se dispone del siguiente cuadro:

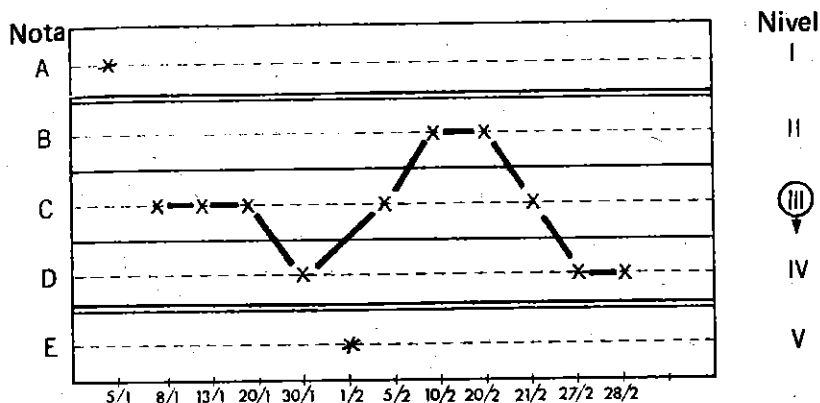
Nota		Nivel
A	<hr/> <hr style="border-top: 1px dashed;"/>	I
B	<hr/> <hr style="border-top: 1px dashed;"/>	II
C	<hr/> <hr style="border-top: 1px dashed;"/>	III
D	<hr/> <hr style="border-top: 1px dashed;"/>	IV
E	<hr/> <hr style="border-top: 1px dashed;"/>	V
Fechas	<hr/>	

Un profesor utiliza tantos cuadros como aspectos desee distinguir en su curso: dictado, gramática, ortografía, redacción...

En la medida en que se van llevandó a cabo las evaluaciones, se hace una cruz en el cuadro y se indica abajo la fecha.

En el momento de hacer la síntesis, se empieza por tachar (¡sin borrarla!) la cruz correspondiente a la nota más elevada y la que representa la más baja, con el fin de eliminar notas que podrían ser accidentales.

Supongamos que se obtiene el siguiente cuadro:



Fechas = 1970

Interpretación

- 1) Los resultados se concentran en tres bandas: II, III, IV. Por lo tanto, no son excelentes, ni flojos o insuficientes. La zona de concentración se marca con trazos dobles.
- 2) El mayor número de notas (el modo) se halla en la banda III. Es la tendencia dominante.
- 3) Como la banda más marcada, después de la III, es la IV, hacemos una flecha en esa dirección, debajo de la III.
- 4) Uniendo las cruces, obtenemos un perfil que señala:
 - a) Una homogeneidad bastante buena de los resultados. Si el alumno dispersara sus resultados anárquicamente

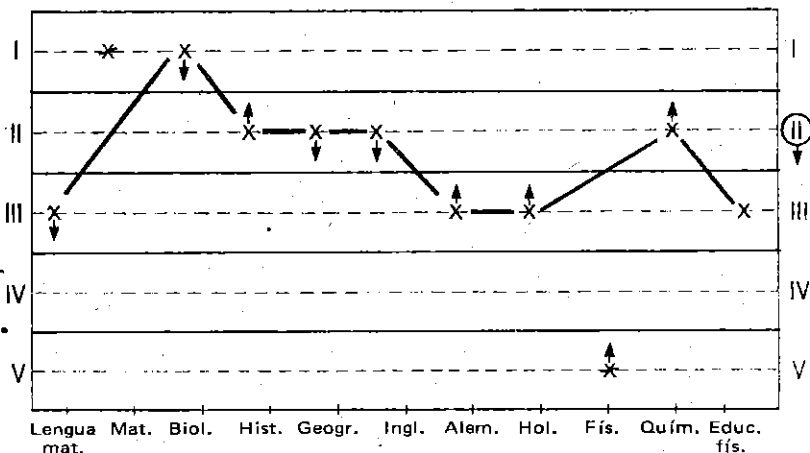
sobre las cinco bandas, habría que aclarar si se halla él solo en ese caso (causa por determinar a nivel del alumno), o si la mayoría de la división se encuentra en igual situación (se ha de ver si los trabajos del profesor no están mal ajustados).

- b) Un descenso al final del período. No está muy marcado: puede explicarse por el cansancio, una enfermedad benigna. . .

Según vemos, no basta con tomar nota de una situación indicada por el perfil; hay que tratar de descubrir la causa. De esta manera, los comentarios orales o escritos de los maestros esclarecen y guían al alumno y a sus padres.

Cabe señalar, por lo demás, que ese sistema no puede funcionar si el profesor se limita a dos o tres evaluaciones, y menos aún con una sola. Esta vez, la introducción de una nueva forma de evaluación mejorará positivamente el comportamiento de ciertos maestros abrumados por la cantidad de material por enseñar. . . o alérgicos a la larga tarea de calificar los trabajos.

CUADRO PARA EL CONJUNTO DE MATERIAS PERIODO DEL AL 19. .



Materias

Interpretación

Cuatro notas de conjunto se hallan en la banda II y cuatro en la III. La tendencia del perfil (aquí hacia lo alto) decidirá respecto de la dominante: en este caso la banda II con flecha hacia abajo.

En caso de que dos niveles sean prácticamente iguales, se puede decidir qué el nivel superior es el que prevalece (sobre todo si se trata de dos niveles contiguos), o bien encerrar dos números y unir los dos círculos con una línea.

Observaciones

1. Este sistema de síntesis de las evaluaciones no posee ningún poder mágico. Si se cometen los errores señalados en este capítulo, la nota de síntesis carecerá probablemente de sentido.

Un fenómeno de concentración en las bandas del medio, muchas veces causa extrañeza a quienes emplean el sistema. Olvidan que en el sistema tradicional, cuyo resultado es un porcentaje, las notas se concentraban, también, en la misma zona: entre el 65% y el 75%, o bien el 80%, mientras que todo el resto de la escala (de 100 puntos) quedaba muchas veces casi vacío.

En realidad, el fenómeno de concentración salta más a la vista porque la escala de cinco grados está mucho mejor visualizada que la antigua.

2. Nada impide ponderar las notas, es decir, acordar, por ejemplo, dos veces más importancia al trabajo en clase que al hecho en casa, a tal aspecto de una materia más que a otro, etcétera.
3. *En la parte dedicada a la "moderación" se encontrarán indicaciones suplementarias para explotar las notas, en particular para determinar los resultados y las calificaciones a fin de año (véase pág. 181).*

G. Un caso particular: la calificación de la redacción francesa

Este problema merece algunas consideraciones especiales porque figura entre los más difíciles.

Si fuese resuelto, la docimología lograría tal vez su más bella victoria. Pero por desgracia, o más bien, felizmente para

la él
del
a en
esor

muy
ifer-

ción
De
tros

mede
s, y
una
oor-
de
icar

I
II
III
IV
V

el hombre, la más noble de sus actividades, la evaluación de lo bello, lo verdadero y lo bueno, se sustraerá siempre, sin duda, a la cuantificación objetiva y, por consiguiente, susceptible de automatización.

1) *Cuatro métodos de evaluación.*

Los principales métodos se reducen a cuatro: el método de la impresión general, la escala de modelos, el método analítico y el de los recuentos de frecuencias.

a) El método de la impresión general

Es el más subjetivo. Se le reprocha la falta de fidelidad en un mismo corrector, los desacuerdos entre varios encargados de la calificación, especialistas o no, y su frecuente ineficacia: más de un profesor ha gastado sus energías en calificar conscientemente centenares de trabajos por año sin obtener resultados convincentes.

Mucho podría decirse acerca de la metodología de la redacción francesa y, sobre todo, de la necesidad de una profunda reforma de los hábitos escolares que consideran a cada alumno como un escritor en potencia; pero no nos corresponde entrar en el tema. Tengamos presente, simplemente, que la conjugación de dos deficiencias, a menudo marcadas, una docimológica y otra metodológica, grava el ejercicio de redacción con una hipoteca tan pesada que algunos autores lo declaran inútil. No compartimos esa opinión.

A pesar de sus deficiencias, el método de la impresión general es el que se halla más en armonía con la complejidad esencial de la redacción. Mediante la utilización de una escala breve de tres o, a lo sumo, de cinco grados, se logra una coincidencia bastante alta entre los que poseen experiencia en la calificación y son de igual formación y de un mismo medio escolar. Veremos más adelante qué precauciones se tomarán antes de declarar un fracaso.

Creemos que una composición corregida mediante ese método no debe aspirar a mejorar mucho el estilo o la simple corrección del lenguaje escrito; existen otros ejercicios para tal fin.

La composición tradicional ha de ser una tarea excepcional, destinada a descubrir talentos o a controlar su maduración y, además, a comprobar debilidades, carencias, impericias,

que se combatirán con ejercicios sistemáticos de enriquecimiento y corrección.

Si se decide atenerse únicamente a la apreciación global, la *técnica de evaluación con limitación de tiempo* merece nuestra atención.

Existen algunas variantes menores. El método propuesto por S. Wiseman⁸ es característico. Cuatro personas encargadas de calificar evalúan, en forma independiente, trabajos hechos en treinta minutos, con un ritmo aproximado de cincuenta redacciones por hora (lapso impuesto para obligarlos a decirse muy rápidamente). Luego se calcula la media aritmética de las notas.

Ese método es probablemente el más eficaz para los trabajos narrativos o descriptivos. En cuanto a las "disertaciones", denominación que abarca todas las redacciones en que el autor debe argumentar, la superioridad del método es más discutida. Sin embargo, parece también válido en ese campo si quienes califican son de un alto nivel intelectual.

La validez general del método y su aplicabilidad a la "disertación" fueron confirmadas por dos investigaciones experimentales de J. Britton.

1ª experiencia⁹

Un conjunto de redacciones escritas por 168 alumnos de 15 años fue calificado, en primer término, según el método analítico, por un experimentado examinador. Después, los mismos trabajos fueron evaluados en forma independiente, según el método de la impresión general con limitación de tiempo, por ocho personas encargadas de calificar.

Con el fin de disponer de un criterio, cinco trabajos más de cada alumno fueron calificados por dos examinadores.

Las notas analíticas y la nota media de tres impresiones rápidas extraídas al azar de entre las ocho de que se disponía

⁸ S. Wiseman, "The Marking of English Composition in Grammar School Selection", en *British Journal of Educational Psychology*, XIX, nov. 1949, pág. 208.

⁹ J. Britton, "Experimental Marking of English Composition Written by Fifteen-Year-Olds", en *Educational Review* (Birmingham), vol. 16, 1, 1963, págs. 17-23.

para cada estudiante, se compararon entonces con las notas que servían de criterio:

correlación entre notas analíticas y criterio: 0,71

correlación entre notas rápidas y criterio: 0,76

Además, la fidelidad de la calificación rápida se calculó mediante la comparación de las notas de dos grupos de tres examinadores, elegidos al azar entre los ocho. Se obtuvo un coeficiente muy elevado: 0,87.

2ª experiencia¹⁰.

Esa segunda experiencia, mucho más detallada que la anterior, abarcó una muestra de 500 alumnos de 17 años, estratificada según el sexo, el tipo de colegio secundario y el lugar de residencia (Londres centro y suburbios, ciudades grandes y pequeñas.) Estaban representados, además, todos los tipos de temas de redacción, desde la simple narración hasta la disertación filosófica y la amplificación poética.

Una vez, más, la calificación múltiple según la impresión rápida se mostró más fiel y válida que la analítica.

b) Las escalas de modelos

Un reducido número de redacciones, a menudo cinco, se eligen a fin de constituir modelos representativos de los distintos niveles de calidad: de lo mediocre a la excelente.

La persona que califica evalúa los trabajos por comparación con los cinco textos patrones de la escala.

Después de gozar de un auge considerable en ciertos países¹¹, las escalas de modelos fueron abandonadas por varias razones¹²:

1. Es raro que la redacción por calificar se parezca de una manera estrecha a los textos patrones.

¹⁰ J. Britton, N. Martin y H. Rosen, *Multiple Marking of English Composition*, Londres, H. M. S. O., 1966.

¹¹ Para un estudio de conjunto, véase: E. Hinton, *Study of the Qualities of Style and Rhetoric Found in English Compositions*, New York, 1940.

¹² R. Braddock y colab., *Research in Written Composition*, Champaign, Ill., N. C. T. E., 1963, pág. 12.

2. Quienes califican tienden a reparar tan solo en las características comunes que les interesan de un modo especial y se desentienden del resto.
3. Se requieren escalas diferentes según el tipo de tema y el nivel de los alumnos.

D. Pidgeon señaló que una extensa experiencia, llevada a cabo hace poco en Gran Bretaña con una escala de modelos de 5 grados, dejó como saldo una diferencia media de 2 grados entre los encargados de la corrección. El resultado, en consecuencia, apenas si es mejor que en condiciones puramente subjetivas.

Cabe observar, sin embargo, que por lo menos a partir de 1903 el método de las escalas de modelos vuelve periódicamente a retener la atención de los investigadores.

c) El método analítico

Dos o tres de las personas encargadas de la calificación se ponen de acuerdo respecto de cierto número de puntos para cada una de las diversas cualidades que se ha convenido en observar. Se determina el total y se calcula la media entre los encargados de la calificación.

Este método es el más difícil de los cuatro. Se lo critica a menudo porque disecciona un todo que, precisamente, parece escapar a toda disección sistemática. S. Wiseman observa, además, que en un conjunto calificado en forma analítica, el trabajo estimado unánimemente como el mejor por los expertos, pocas veces se clasifica en primer lugar.

Sin embargo, la investigación muestra que, si las cualidades por observar y la importancia por atribuirles se definen con precisión, el método analítico es el más seguro si debe confiarse en una sola persona para corregir. El método de la impresión general también proporciona buenos resultados, como ya vimos, pero exige la participación de varias personas encargadas de la calificación¹³.

¹³ Se hallará, asimismo, un buen estudio comparado en B. Cast, "The Efficiency of Different Methods of Marking English Composition", en *British Journal of Educational Psychology*, IX, nov. 1939, págs. 257-269, y X, feb. 1940, págs. 49-60, citado por R. Braddock, *op. cit.*, pág. 13.

Para ser eficaz y practicable, el método analítico tiene que cumplir, aparentemente, con dos exigencias mínimas:

- 1° Las personas que corrigen deben aceptar el plan de análisis. Lo ideal es que participen en su elaboración.
- 2° El plan no ha de resultar demasiado detallado. De esta manera, el método muy sutil utilizado por E. Burton¹⁴ solo es interesante para la investigación, pues requiere un tiempo excesivo.

¿Qué cualidades se observarán?

C. Remondino mostró¹⁵ que las actitudes fundamentales de los encargados de la corrección (profesores o no), poseedores de cierta cultura, no difieren esencialmente: las mismas cualidades retienen su atención. Pero en la apreciación de conjunto les atribuyen pesos muy diversos, y es ahí donde se halla, en realidad, la fuente de las divergencias.

Elaborar un cuadro muy detallado de análisis del contenido es un engaño, pues aun existiendo matices determinados, la gente que califica los vuelve a mezclar inconscientemente. El análisis factorial pone en evidencia ese proceso. El trabajo de Remondino ofrece interesantes resultados en ese sentido.

En primer término, este autor interrogó largamente a veinte profesores de materias literarias (nivel secundario) y elaboró la lista de las cualidades estudiadas en las redacciones escolares. De esta manera compone el cuadro de la página siguiente:

¹⁴ E. Buxton, "An Experiment to Test the Effects of Writing Frequency"... en *Alberts Journal of Educational Research*, V., jun. 1959, págs. 91-99

¹⁵ Estudio factorial sobre la calificación de las redacciones escolares referentes a la lengua materna, en *Le Travail Humain*, XXII, en. jun. 1959, págs. 27-40.

1. Legibilidad	Cualidad de un tipo de escritura que se presta para una lectura fácil, rápida e inequívoca.
2. Estética	Trazo armonioso y agradable de las letras y buen gusto en cuanto a su distribución en la hoja.
3. Presentación	Limpieza, cuidado, buena presentación del trabajo.
4. Exactitud de la ortografía	Densidad de los errores de ortografía.
5. Exactitud morfológica	Densidad de los errores morfológicos.
6. Exactitud sintáctica	Densidad de los errores de sintaxis.
7. Estructura de la exposición	Calidad de una exposición ordenada, en las proporciones propuestas y de acuerdo con un plan.
8. Riqueza de ideas	Cantidad de ideas; recursos utilizados.
9. Pertinencia de las ideas	Cualidad según la cual las ideas expuestas se hallan en justa relación con el tema tratado.
10. Precisión de información	Veracidad y exactitud objetiva de las afirmaciones y los hechos expuestos.
11. Exhaustividad	Cualidad que consiste en no omitir nada de todo lo que debe decirse.
12. Concisión	Cualidad según la cual las cosas se expresan con el mínimo indispensable de términos, sin repeticiones, redundancia ni tortuosidad.
13. Propiedad del lenguaje	Empleo justo de los términos.
14. Estilo	Facilidad, exactitud, dominio del lenguaje desde el punto de vista de la "construcción de oraciones".
15. Originalidad	Cualidad mediante la cual se trasluce y se afirma en el trabajo algún rasgo de la personalidad.
16. Madurez	Capacidad de juicio, profundidad crítica, agudeza de los razonamientos expuestos.
17. Imaginación	Capacidad de creación, de trasfiguración, de "proyección" revelada por el trabajo.

Un análisis factorial mostró entonces que esas 17 cualidades dependían únicamente de cuatro grupos:

1. Presentación gráfica (1, 2 y 3).
2. Uso del lenguaje (4, 5, 6, 13 y 14).
3. Contenido y organización de la exposición (7, 8, 9, 10 y 11).
4. Aspectos personales del fondo (15, 16 y 17).

Un ejemplo

Las conclusiones de Remondino confirman ampliamente las de los otros investigadores. Las siguientes escalas, construidas por el *Educational Testing Service* (E.T.S.) de Princeton¹⁶, y levemente retocadas por E. Page¹⁷, dan prueba de la similitud de propósitos.

Traducimos el texto completo porque constituye uno de los mejores ejemplos que es factible ofrecer actualmente para indicar el camino por seguir.

CRITERIOS PARA CALIFICAR LAS REDACCIONES

I. Definición de los rasgos por evaluar.

- A. Ideas o contenido: la cantidad y calidad del material utilizado para tratar el tema.
- B. Organización: la relación entre las partes de la redacción y el total.
- C. Estilo: uso del lenguaje más allá de la simple corrección gramatical.
- D. Mecánica: ortografía, gramática, puntuación.
- E. Creatividad.

II. Guía para la evaluación de estos cinco rasgos

A. *Ideas o contenido*

Nivel elevado

El estudiante escribe sobre todos los puntos señalados por el tema o el plan de trabajo. Comprende bien el tema y utiliza

¹⁶ E. T. S., *Definitions of Ratings on the E. T. S. Compositions Scale*, citado por E. Page, *op. cit.*, págs. 76-77.

¹⁷ *Op. cit.*, págs. 78-80.

definiciones claras. Sabe considerar el tema con una perspectiva más amplia que la de sus compañeros. Con otras palabras, da pruebas de una experiencia más rica.

Nivel medio

Las ideas son apropiadas, pero convencionales y poco numerosas. Ciertos aspectos del tema se pasan por alto. El alumno no parece estar dotado de una mente demasiado rica en ideas.

Nivel bajo

El estudiante omite muchos aspectos importantes del tema. No dispone, aparentemente, de una reserva de conocimientos correspondientes a ese tema y, por consiguiente, repite en forma continua algunas ideas simples.

B. Organización

Nivel elevado

El alumno sigue un plan definido. Si presenta el pro y el contra, alega razones pertinentes en un orden eficaz. Si describe algo, lo hace de una manera ordenada (del vértice a la base, por orden de importancia, de complejidad, etcétera). Si explica un concepto o proceso, emplea un plan coherente de análisis, de definición o de ilustración. Advierte con claridad aquello que hace referencia a su plan y evita repeticiones. Muestra el sentido de la medida al tratar las diferentes partes de su trabajo.

Nivel medio

El estudiante no se atiene a su plan o introduce ideas sin relación con el tema. Dedicar demasiado tiempo a cosas de escasa importancia o se repite. Trata el tema por libre asociación (¿qué es lo que se me ocurre cuando pienso en Hawaii?) en vez de perseguir un fin definido.

Nivel bajo

El alumno no se ha preguntado aparentemente qué iba a hacer, antes de empezar a escribir. No sigue plan alguno. El trabajo se encamina en una dirección, luego la cambia una y

otra vez hasta que el lector se siente perdido. Los puntos principales no están separados de una manera clara unos de otros y su orden de presentación queda librado al azar.

C. *Estilo*

(Varios aspectos del estilo pueden intervenir en la evaluación: individualidad, vivacidad, elegancia, etcétera. Sin embargo, aquí nos interesan tan solo tres aspectos estilísticos: claridad, variación y despliegue de recursos lingüísticos.)

Nivel elevado

El estudiante emplea un lenguaje que facilita la comprensión del trabajo. Utiliza términos adecuados en su sentido habitual y los presenta en un orden normal. Maneja bien las transiciones. El alumno evita las ambigüedades y no engaña la expectativa del lector. Al mismo tiempo elude la repetición monótona de palabras, complementos o estructuras de oraciones. Por último, da pruebas de poseer un gran caudal de recursos lingüísticos. Su vocabulario es bueno. Utiliza estructuras paralelas o hace un empleo sutil de la subordinación.

Nivel medio

A veces, el estudiante desorienta al lector utilizando una palabra inadecuada, un giro estafalario, una metáfora poco clara, desplazando de manera inoportuna un complemento o una subordinada o haciendo transiciones bruscas. La repetición de palabras, giros y estructuras se vuelve monótona. Los recursos lingüísticos son imitados. Gusta de clisés y giros rebuscados.

Nivel bajo

Utiliza las palabras de una manera vaga. Giros ambiguos, construcciones endebles, vocabulario y estructura de frases infantiles.

D. *Mecánica*

Nivel elevado

La estructura de las oraciones resulta por lo general correcta, aun cuando se trata de modelos variados y complica-

los. Se respetan las reglas de ortografía, e incluso, las palabras difíciles se escriben comúnmente sin error. No hay violación grave de las reglas de puntuación, mayúsculas, abreviaturas y escrituras de números.

Nivel medio

Ocasionales faltas de sintaxis. A veces mala ortografía de palabras difíciles. Alguna que otra violación de las reglas de puntuación, etcétera.

Nivel bajo

Muchísimos errores.

E. *Creatividad*

Nivel elevado

El estudiante sorprende con formas nuevas y eficientes de considerar el problema. Introduce ideas nuevas en su manera de tratar el tema. Halla formas frescas e interesantes de utilizar el lenguaje a fin de destacar sus ideas.

Nivel medio

El estudiante piensa en aquello que se esperaba que pensara. Trata las cosas más o menos como todo el mundo. Emplea expresiones y estructuras de oraciones comunes.

Nivel bajo

El alumno emplea clisés de pensamiento y de expresión. Desarrolla el tema en forma superficial. Repite fórmulas sin comprender realmente su significado.

d) El método de los recuentos de frecuencias

Aunque no pertenece a la categoría de las evaluaciones subjetivas, lo presentamos aquí para completar el cuadro de métodos.

Consiste en definir tipos de errores (después de un estudio de diagnóstico), en contarlos en la redacción y en expresar el número encontrado cada 100 o 1.000 palabras.

En vez de adoptar un punto de vista negativo, existe la posibilidad, asimismo, de hacer el inventario (relativamente

exhaustivo) de los distintos procesos utilizados en la redacción. El mejor ejemplo de tal método se halla en *Eléments pour servir à l'étude d'analyses littéraires*, de A. Purves¹⁸.

Un último método positivo consiste en aislar ciertos aspectos que, en un umbral determinado, se revelan "predictivos" de la nota que la gente experimentada en la corrección suele asignar. En ese caso, una computadora, programada adecuadamente, se encarga de la calificación. De esta manera obró E. Page.

Es probable que los recuentos de frecuencias ofrezcan la única esperanza de una calificación objetiva de las redacciones. Sin embargo, aparentemente, deben permanecer dentro del dominio de la investigación ajena a la práctica escolar, a menos que la computadora llegue a ser tan accesible que el estudio de las frecuencias se realice de una manera económica y constituya uno de los aspectos de la evaluación. Esto no se excluye de manera alguna.

2) ¿Varios temas para elegir?

J. Britton observa acertadamente que la redacción de examen sobre un tema único implica, a menudo, que el profesor sostenga la hipótesis de que un tema como "Nubes" revelará la aptitud del estudiante en la misma medida que otro titulado "Exponga sus opiniones acerca de la pena capital".

Se ha demostrado experimentalmente que no es así; volveremos aún sobre este punto. No obstante, es dudoso que la justicia quede mejor resguardada si se ofrecen tres temas para elegir, puesto que el puntaje se atribuye a una sola de las redacciones. En primer término habría que demostrar que con tres temas, e incluso con cinco, se cubre todo el espectro de estímulos imprescindibles para otorgar a cada uno la misma oportunidad, y aunque esto se confirmara, sería importante todavía que el profesor escogiera siempre los temas con entero conocimiento de causa.

Afirmamos que la experiencia nos ha demostrado que el rendimiento del alumno varía según el tema propuesto. Finlay-

¹⁸ A. Purves y V. Rippère, "Eléments pour servir à l'étude d'analyses littéraires" (traducido del inglés por J. Dubois), en *Scientia Paedagogica Experimentalis*, VI, 2, 1969, págs. 229-334.

son¹⁹ comprobó, con respecto al período final de la escuela primaria, que si se invita a los alumnos, con un intervalo de una semana, a hacer una redacción sobre la base de una misma elección de cuatro temas, los que cambian de tema (más de la mitad en el experimento) obtienen notas significativamente diferentes.

Wiseman y Wrigley²⁰ llegan a la misma conclusión (elección entre el mismo conjunto de cinco temas con cuatro meses de intervalo) y demuestran que la diferencia de nota no se debe a una falta de fidelidad del profesor. Vernon y Mallican²¹ confirman esto a nivel universitario.

¿Qué hacer? Si queremos verificar la capacidad de expresarse por escrito desde un punto de vista estrechamente definido (por ejemplo: capacidad para redactar un informe científico o una carta comercial), la sensibilidad particular de los candidatos apenas si entra en consideración, y un solo tema es suficiente.

En cambio, si se trata de descubrir una aptitud para la expresión en cualquier terreno, o bien de conducir a los alumnos a discutir problemas en los cuales la disposición espiritual y el nivel de información pueden desempeñar un papel importante, ya no es factible emitir un juicio en función de una sola redacción. Entonces, lo ideal consiste en referirse al trabajo de todo el año. Si el alumno pudo elegir, en cada oportunidad, entre varios temas (cinco, por ejemplo), en pocos meses habrá reaccionado ante varias decenas de temas, y cabe suponer que cada uno habrá tenido realmente una oportunidad. Sería tentador dejar a la inventiva personal la elección del tema, si fuese posible tomar precauciones contra la preparación fraudulenta. Y esto es muy difícil. . . .

¹⁹ D. S. Finlayson, "The Reliability of the Marking of Essays", en *British Journal of Educational Psychology*, XXI, 2, 1951, citado por Britton (1966).

²⁰ S. Wiseman y J. Wrigley, "Essay-Reliability: the Effect of Choice of Essay-Title", en *Educational and Psychological Measurement*, 18, 1, 1958.

²¹ P. Vernon y G. Millican, "A Further Study of the Reliability of English Essays", en *British Journal of Statistical Psychology*, VII, 2, 1954.

3) *Conclusión*

El destino que se reservará a la redacción francesa dependerá del fin que se persiga. ¿Qué queremos? ¿Corregir un defecto particular? En tal caso, mejor será examinar ese único aspecto en los trabajos o abordarlo con ejercicios especiales.

¿Conocer la capacidad de los alumnos para corregirse, para identificar la prosa de acuerdo con el uso correcto? En ese caso, un *test* objetivo, con respuestas de elección múltiple, puede alcanzar.

¿Lograr la adquisición de la capacidad de expresarse con claridad? Entonces, temas muy limitados permiten concentrar la atención sobre el objeto del trabajo. ¿Por qué habría que excluir temas tales como "Describa una bicicleta de carrera"?

¿Descubrir talentos de escritores? Hay que dejar libertad a la expresión del alumno e invitarlo, a menudo, a que seleccione su tema.

Pero no nos engañemos: son diferentes las capacidades que de esta manera se ponen a prueba. Con el segundo experimento de J. Britton, anteriormente descrito, el autor ha demostrado:

- 1° Que la correlación entre los resultados de redacción creativa y un resumen de texto varía de 0,30 a 0,40.
- 2° Que la correlación de la redacción con un *test* de comprensión de texto resulta un poco más elevada: de 0,35 a 0,45.

Los desniveles son, pues, considerables.

Como ya hemos expresado, no opinamos que deba eliminarse la redacción tradicional. En algunos países, en los que se la había suprimido por la dificultad de una evaluación rigurosa, se la volvió a introducir, sobre todo en Estados Unidos, donde la Comisión de Exámenes de Admisión en la Enseñanza Superior renunció al "*Test* objetivo de redacción", y en Inglaterra, que hizo lo propio con las dos pruebas clave de selección: los exámenes 11 + y 16 +.

No obstante, si la redacción sigue desempeñando un papel importante en los exámenes, debe modificarse la manera de calificar, aún demasiado difundida. Hacemos nuestra la proposición de R. Braddock:

"Si hay que evaluar un gran número de estudiantes para decidir acerca de su éxito o su fracaso, es importante permitir a los que se malogran hacer un segundo trabajo . . . En caso de

asignar grados, se requieren por lo menos tres trabajos; los dos mejores serán conservados y se calculará la media . . . Los trabajos han de ser calificados por dos personas, como mínimo, y éstas emplearán un sistema de calificación claramente comprendido y aceptado, en cuya aplicación estén bien ejercitadas.”²²

III. La calificación objetiva

La calificación objetiva prescinde de la opinión personal de quienes califican. El ejemplo más sencillo sería el de un problema que comprende diez multiplicaciones, a razón de un punto por cada resultado correcto. Asimismo, la calificación de un *test* compuesto de ítem de elección múltiple se realiza mediante el simple recuento del número de elecciones correctas.

De una manera más general, se denomina calificación objetiva a “la asignación de valores numéricos a muestras de comportamiento suficientemente limitadas, definidas y controladas, a fin de permitir un acuerdo general entre los jueces o las personas que califican”.²³

Tal como ya observamos, la objetividad de la calificación no garantiza de manera alguna la objetividad del examen en su conjunto. En última instancia, la elección de las preguntas es siempre subjetiva, y resulta fácil imaginar un examen en el que ninguno de los problemas ofreciera una muestra verdadera de la capacidad que se pretende evaluar. Por objetiva que sea la corrección, en un caso así, la prueba carecería de validez. Abundan los ejemplos de semejante desencuentro.

IV. Estandarización o medida de la posición relativa

La interpretación de cualquier nota plantea un problema de “comparabilidad”.

Estandarizar consiste en definir *normas*, es decir, distribuciones estadísticas que permitan ubicar los rendimientos individuales.

²² R. Braddock y colab., *op. cit.*, pág. 45.

²³ J. C. Flanagan, “Units, Scores and Norms, en E. Lindquist, Ed., *Educational Measurement*, Washington, A. C. E., 1961, 4a. edición.

Según el fin que se persiga, las normas son establecidas por un grupo limitado o por conjuntos más numerosos que pueden abarcar hasta la población de un país. En la práctica, su validez se limita siempre a un conjunto determinado, y para utilizarlas es importante saber de cuál se trata.

Durante mucho tiempo, el porcentaje ha sido el alfa y el omega de la norma escolar. De esta manera, para estandarizar algunos tests no se hacía más que expresar en porcentaje el número de respuestas correctas. Las deficiencias de semejante sistema son bien conocidas.

En la actualidad desempeña un papel determinante la diferencia entre individuos y no la posición en una escala del saber considerada como absoluta²⁴. La media o la mediana²⁵ constituyen los puntos de referencia a partir de los cuales se construyen escalas "universales" que, dentro de ciertas condiciones sobre las cuales aún hablaremos, permiten expresar los puntajes brutos, obtenidos en distintas pruebas, en unidades comparables.

Los sistemas de estandarización más conocidos y accesibles son los siguientes:

1. Los percentiles.
2. Las notas estándar o notas Z.
3. La escala normalizada de cinco clases.
4. La escala normalizada de 9 clases (Stanines)*.

Pasaremos a estudiarlos sucesivamente.

A. Los percentiles.

1) Definición

El percentil o rango ocupado en una escala de cien grados goza aún de cierto favor a causa de su parecido exterior con el porcentaje, del que difiere, no obstante, de manera fundamental. El porcentaje indica qué proporción del total de los puntos

²⁴ Véase naturaleza y deficiencia de las escalas de evaluación. Calcular el porcentaje equivale a considerar una escala de cien grados.

²⁵ Recordemos que la mediana es el punto que divide una serie de notas ordenadas en dos partes iguales: 1 - 2 - 3 - 4 - 5 6 1 - 2 - 3 - 4 - 5 - 6.

*Stanines: Designación de grados en una escala contrastada en nueve puntos (*standard nines*) (N. del T.).

asignados a un examen obtuvo el alumno. El percentil indica cuántos alumnos se clasificarían después de un alumno determinado, si la división contara con cien estudiantes. Por lo tanto, un puntaje equivalente al 90° percentil es superior al 90% de la población considerada.

La mediana corresponde al 50° percentil.

2) Cálculo

En principio, no se calculan percentiles si el número de notas es inferior a cien. El siguiente ejemplo se basa en las cien notas que ya nos sirvieron al estudiar la curva de Gauss.

Procedimiento general:

- Confeccionar el cuadro de clases (empezando por la clase inferior).
- Indicar las frecuencias (f).
- Calcular las frecuencias acumuladas (f_c): cada frecuencia se suma a la anterior.
- Aplicar la fórmula:

$$C_x = I + \frac{(N \cdot C)/100 - f_c}{f_s} \times i$$

C_x = valor de la nota (x) en percentil.

$(N \cdot C)/100$ = el rango de la nota correspondiente a un percentil cualquiera.

Ejemplo: el 50° percentil corresponde a la nota mediana, o sea, $N/2$. En cuanto al resto, se trata de una simple regla de tres. El 75° percentil será, pues: $(100 \times 75) : 100 = 75$ (los números son los mismos porque, en nuestro ejemplo, hay precisamente 100 notas).

I = la media entre el límite superior de la clase en que se halla el percentil buscado y el límite inferior de la clase anterior.

f_c = frecuencia acumulada de la clase que precede a aquella en la que se encuentra el percentil buscado.

f_s = frecuencia simple del intervalo donde se encuentra el percentil buscado.

i = intervalo de clase.

Ejemplo:

	Clases	Frecuencias (<i>f</i>)	Frecuencias acumuladas (<i>f_c</i>)
	24-28	1	1
	29-33	2	3
	34-38	8	11
	39-43	10	21
	44-48	23	44
	49-53	22	66
	54-58	18	84
	59-63	9	93
	64-68	5	98
	69-73	1	99
	74-78	1	100

	Cálculo del 50° percentil	Cálculo del 16° percentil	Cálculo del 84° percentil
<i>N.C</i>	100×50	100×16	100×84
<i>N</i>	100	100	100
<i>l</i>	$\frac{49+48}{2} = 48,5$	$\frac{39+38}{3} = 38,5$	$\frac{54+53}{2} = 53,5$
<i>f₁</i>	44	11	66
<i>f₂</i>	22	10	18
<i>C.</i>	$48,5 + \frac{50-44}{22} \times 5 = 49,9$	$38,5 + \frac{16-11}{10} \times 5 = 41$	$53,5 + \frac{84-66}{18} \times 5 = 58,5$

N.B. Decilaje: el 1er. decil, etcétera.

*Diferencia semi-intercuartil: (75° C - 25° C)/2.*3) *Crítica*

La escala de percentiles es, como ya expresamos, una amplia escala de evaluación de 100 grados. Informa acerca del rango ocupado, pero no sobre la distancia entre rangos. Ahora bien, si la distribución es normal, esa distancia va disminuyendo a medida que nos acercamos a la mediana (igual a la media en caso de normalidad perfecta).

El siguiente cuadro muestra con toda claridad ese fenómeno.

	-2σ	-1σ	0	$+1\sigma$	$+2\sigma$
Percentiles	2°	16°	50°	84°	98°

Comprobamos que 68 percentiles se hallan aglomerados alrededor de la media; esto ofrece una clasificación demasiado burda para la mayoría del grupo. Por ejemplo, se corre el riesgo de considerar que existe una diferencia importante entre el 75° percentil y el 25°. La antigua norma del porcentaje está aferrada en las mentes y hace pensar que el percentil 75 vale tres veces más que el 25. El cuadro anterior nos muestra que, en realidad, se trata de dos resultados relativamente cercanos y medianos.

De hecho, si se siguen utilizando las normas en función de percentiles, sobre todo para la selección en ciertas universidades estadounidenses, es porque éstas solo seleccionan en la banda de los diez o veinte percentiles superiores, zona donde la discriminación resulta satisfactoria.

B. Las notas estándar o notas Z

Un alumno obtuvo las siguientes notas brutas²⁶. Cálculo: 22 sobre 25; lectura: 72 sobre 100; ciencias: 26 sobre 50.

El examen de esas notas no nos enseña gran cosa. Si el profesor es severo en lectura y mucho menos en cálculo, el 72/100 en lectura será tal vez más meritorio que el 22/25 en cálculo. Además, las notas brutas no permiten la comparación con otros alumnos de un mismo nivel pedagógico, por ejemplo.

Con el fin de posibilitar la comparación, esas notas se expresan en función de los desvíos tipo; esto permite ubicarlas sobre una misma curva (notas estandarizadas o puntajes estándar).

²⁶ Adaptado según R. Thomas, *op. cit.*

Supongamos que obtenemos los siguientes resultados:

	Media	Desvío tipo
Cálculo	15	2,5
Lectura	50	10
Ciencias	29	5

Cálculo de los puntajes estándar: $z = (x - M)/\sigma$

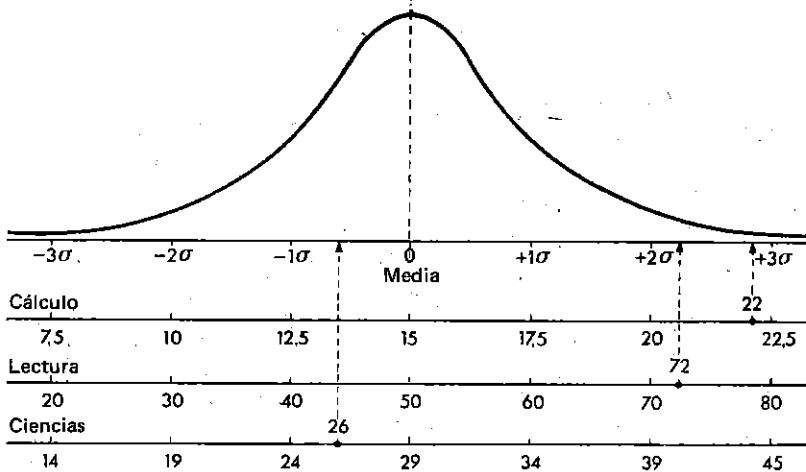
Ejemplo: $(22 - 15) : 2,5 = 2,8 \sigma$.

Conociendo z , podemos volver a encontrar la nota bruta:

$$x = M + z \sigma$$

$$x = 15 + (2,8 \times 2,5) = 22.$$

Representación gráfica:



Aplicación

Los estudiantes I y II se sometieron a una serie de *tests*. Comparar los resultados²⁷.

TEST	Media M	Desvío tipo σ	X		x		Puntajes estándares z	
			Puntajes brutos		Desviacio- nes (M-X)			
			I	II	I	II	I	II
Inglés	155,7	26,4	195	162	+39,3	+ 6,3	+1,49	+0,24
Lectura	37,7	8,2	20	54	-13,7	+20,3	-1,67	+2,48
Información	54,5	9,2	39	72	-15,5	+17,5	-1,67	+1,88
Aptitudes escolares	87,1	25,8	139	84	+51,9	- 3,1	+2,01	-0,12
Atención	24,8	6,8	41	25	+16,2	+ 0,2	+2,38	+0,03
Totales			434	397			+2,54	+4,51
Medias							+0,51	+0,90

Si nos atuviéramos simplemente a los totales (puntajes brutos I y II), el estudiante I sería considerado superior al II. El cálculo de porcentajes confirmaría tal impresión.

Ahora bien, si por medio del cálculo de los puntajes Z reducimos todos los resultados a la misma media y a la misma unidad de desvío con relación con ésta, comprobamos:

- 1° Que con respecto a la media, el estudiante II (0,90) obtiene un resultado casi dos veces más favorable que el I (0,51).
- 2° Que el estudiante II obtiene los resultados más homogéneos.

Crítica

La transformación en puntajes Z se basa en la hipótesis de que, para el grupo considerado, las aptitudes a las cuales se refieren los *tests* se distribuyen de la misma manera y

²⁷ Números tomados de J. P. Guilford, *Fundamental Statistics*, op. cit., pág. 513.

conducen a idénticas medias y dispersiones. Tal hipótesis es prácticamente imposible de verificar, pero, tal como observa J. P. Guilford, podemos estar seguros de que este sistema es, de todos modos, mejor que el cálculo de porcentajes.

Por lo demás, la concentración de la escala alrededor de la media confiere a ese punto una imagen de discriminación errónea. A causa de las inevitables imperfecciones de la calificación, el mero azar hará caer un número más o menos elevado de sujetos a un lado o al otro.

Una transformación

Los puntajes Z toman el cero como media y un desvío tipo igual a uno como unidad.

Algunos autores consideran ese sistema poco elocuente y prefieren expresar los resultados a la manera de los cocientes intelectuales, en los cuales 100 se toma tradicionalmente como media y 145 se considera un cociente de "genio".

La transformación resulta fácil. Se toma cien como media y cada puntaje Z se multiplica por 15.

C. La escala de cinco clases (N entre 50 y 100).

Las clases 2, 3 y 4 implican un desvío tipo.

La 3a. clase se concentra sobre la media.

Las clases 1 y 5 son de extensión ilimitada.

$-1,5\sigma$	$-0,5\sigma$	M \downarrow	$+1,5\sigma$	$+0,5\sigma$
1	2	3	4	5
7%	24%	38%	24%	7%

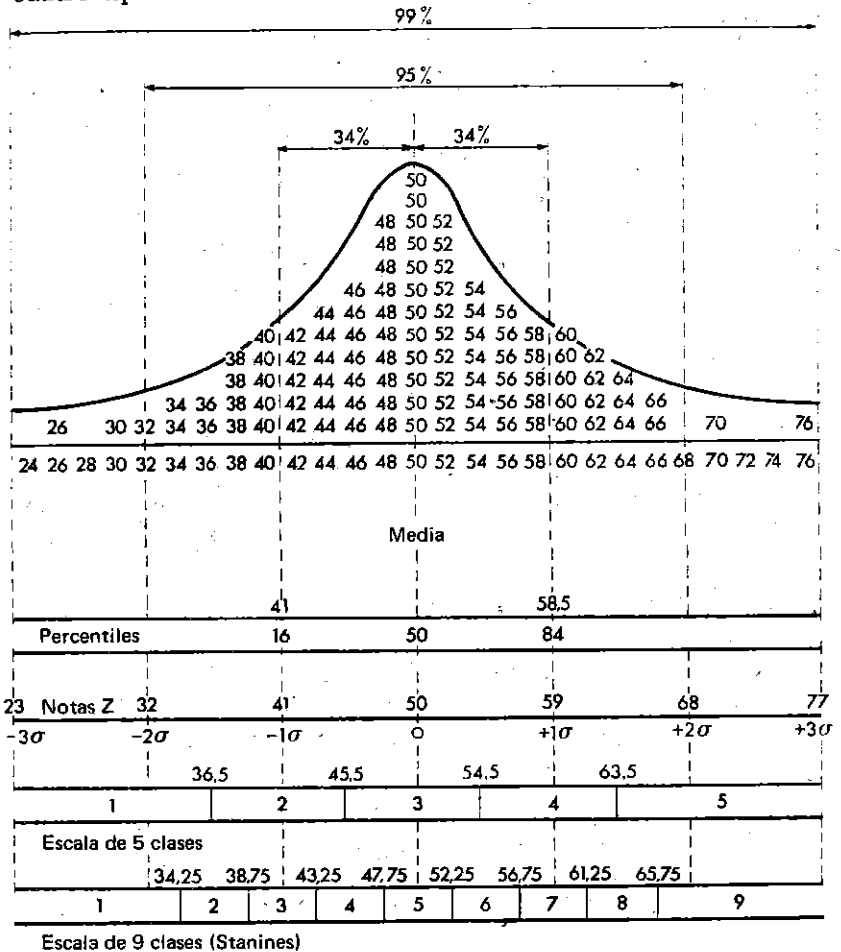
D. La escala de 9 clases (Stanines) (N > 100).

Las clases 1 y 9 son de extensión ilimitada.

Las otras clases admiten un semidesvío tipo; la 5a. se centra en la media.

M								
$-1,75\sigma$	$-1,25\sigma$	$-0,75\sigma$	$-0,25\sigma$	$+0,25\sigma$	$+0,75\sigma$	$+1,25\sigma$	$+1,75\sigma$	
1	2	3	4	5	6	7	8	9
$>$		$>$		$>$		$>$		$>$
4%		7%		12%		17%		20%
$<$		$<$		$<$		$<$		$<$
4%		7%		12%		17%		20%

El siguiente gráfico permite comparar con facilidad los cuatro tipos de normas.



CAPITULO 4

CONTROL DE LA FIDELIDAD DEL EXAMEN

Muchas de las medidas que hemos encarado hasta aquí están enderezadas a asegurar una mayor justicia escolar.

El control de la fidelidad se origina en la misma preocupación: idealmente, un mismo examen (escrito u oral) rendido varias veces, sin que el alumno tenga tiempo de aprender cosas nuevas, debería llevar siempre al mismo resultado. Se desea, pues, que aparezca tan *fiel* como un metro que, bien utilizado, mide siempre la misma longitud con un error mínimo.

Evidentemente, tomar la misma prueba una y otra vez no permite un adecuado control de la fidelidad. La primera prueba provoca una reflexión generadora de aprendizaje. Después, el alumno sigue pensando en el problema, verifica la exactitud de sus respuestas, recoge informaciones complementarias, etcétera. En resumidas cuentas, hay que recurrir a otro medio o, más exactamente, a un conjunto de medidas favorables a la fidelidad.

Si suponemos que se ha logrado la fidelidad de la corrección, queda por garantizar la fidelidad de las respuestas.

1. Evitar toda ambigüedad en las preguntas

Una pregunta que carece de claridad hasta el punto de prestarse a varias interpretaciones diferentes, de entrada priva de fidelidad al examen. En efecto, nada permite prever con alguna certeza cuál de los sentidos hubiera sido otorgado por un mismo individuo o por un grupo en distintas ocasiones.

En la página 74 contemplamos algunos medios destinados a asegurar la claridad de las preguntas.

2. Un número suficiente de preguntas

Si tan solo planteamos unas pocas preguntas, dejamos sin explorar vastas zonas de materias. Ahora bien, puede que el alumno no haya comprendido todas las materias, o bien que haya desatendido partes del curso especulando con la buena suerte o con una preferencia marcada del profesor. Según el área en que incidan las preguntas, obtendremos un resultado brillante o un fracaso.

El único medio para evitar tal aventura consiste en cubrir *toda* la materia. En otras palabras, el examen debe ser un *buen muestreo* del total.

En principio, cuantas más preguntas agregamos, más aumenta la fidelidad.

3. Un control matemático

Cuando se trata de exámenes importantes, deseáramos obtener aún más garantías.

a) *El método pares-impares*

Si el número de preguntas resulta bastante elevado, el siguiente procedimiento simula una situación en la que un mismo alumno rendiría, prácticamente al mismo tiempo, dos veces el mismo examen.

- a) Numeradas las preguntas, se las divide en dos grupos: pares e impares.
- b) Se establecen dos cuadros paralelos de los resultados y se calcula su correlación (r). Expresado en otras palabras, se evalúa numéricamente la relación existente entre los dos grupos¹.
- c) Los dos exámenes artificialmente creados abarcan, cada uno, la mitad del original. En virtud de lo que vimos en el párrafo 2, resultan, por consiguiente, menos fieles, y la correlación subestima la situación real.

¹ El método para calcular la correlación se explica en todos los manuales elementales de estadística.

Una sencilla fórmula permite realizar la corrección necesaria: el índice de fidelidad se obtiene finalmente por
$$\frac{2r}{1+r}$$

Cuanto más se acerque el resultado a 1, tanto más aumentará la fidelidad.

Por lo general se exige 0,80 como mínimo.

Cuando se prepara el examen mediante una *test* previo, una fidelidad insuficiente se trata de corregir, sobre todo, mediante el aumento del número de preguntas.

Si el control de fidelidad se lleva a cabo después del examen, la comprobación de una fidelidad insuficiente inducirá a los maestros a interpretar los resultados con mayor prudencia y a construir los exámenes futuros con más esmero.

b) *Dos formas paralelas*

El control matemático es el mismo que se aplica en el método *pares-impares* pero, en este caso, los examinadores preparan dos series completas de preguntas —consideradas unánimemente como equivalentes— a partir de un mismo tablero de objetivos.

Si la población es bastante numerosa, una mitad, elegida al azar (por ejemplo, según el orden alfabético de los nombres), recibe la primera forma, y la otra, la segunda. Tratándose de poblaciones poco numerosas, se invita a los alumnos a responder a las dos series de preguntas después de algunos días. Esta segunda solución es, evidentemente, mucho menos rigurosa.

4. Repetición de la calificación

La fidelidad de las respuestas no es suficiente. Ha de acompañarse de la fidelidad de la calificación. Esta se controla mediante la invitación a un mismo profesor para que califique dos veces los mismos trabajos, con un intervalo de algunos días o semanas, o bien comparando las notas otorgadas a un mismo trabajo por varias personas encargadas de esa tarea.

CONTROL DE LA VALIDEZ

La fidelidad de un examen no garantiza de manera alguna su validez. Un metro mal construido, que midiera cinco centímetros de más, indicaría la misma longitud en cada medición y, sin embargo, llevaría a una conclusión errónea.

En suma, validar un examen significa probar que mide efectivamente lo que debe medir. Lograr esa certeza es esencial, pues si la conclusión es negativa, todos los esfuerzos desplegados para la organización y el rendimiento de los exámenes son vanos o, por lo menos, están desviados de su objetivo.

Ahora bien, la validación no es cosa fácil. Al hacer el balance de nuestro conocimiento acerca de la validez, R. Cox llega a la conclusión, en 1969, de que la investigación apenas si ha tocado las cuestiones fundamentales¹. Tanto más, entonces, el problema merece la atención y prudencia de los profesores.

Según la índole retrospectiva o prospectiva del examen, se distinguen la validez del contenido y la validez predictiva.

I. La validez del contenido

Se refiere principalmente a los exámenes destinados a establecer el balance de las adquisiciones y, por consiguiente, también de la enseñanza. Se había asignado tal meta y elegido tales objetivos. ¿Fueron alcanzados?

¹ R. Cox, "Reliability and Validity of Examinations", en J. Lauwerys y D. Scanlon, *Examinations, op. cit.*, pág. 43.

En los exámenes referidos a largos períodos de aprendizaje, es evidente que no existe la posibilidad de interrogar en detalle sobre toda la materia. Por esa razón es importante no solo localizar los puntos principales, sino también tener en cuenta su importancia relativa. Si se cumple con estos dos criterios, las preguntas podrán suscitar respuestas *representativas* del conjunto de lo adquirido, de la competencia total.

En otras palabras, la validez del contenido depende de la calidad y adecuación del muestreo de conocimientos y capacidades.

Hemos visto que, en una enseñanza bien concebida, los objetivos por lograr y controlar en el examen han sido definidos al comienzo del año.

Pero en ese momento se trataba de intenciones que las circunstancias pueden haber modificado. Antes de tratar de hacer el balance del trabajo de sus alumnos, el profesor hará, pues, el suyo. Si se omite esa vuelta sobre los propios pasos, la validez de los exámenes es ilusoria o, peor aún, engañosa, destinada a justificar al profesor frente a los ojos de sus superiores. De esta manera se explica que ciertas preguntas se refieran a materias apenas rozadas, pero que figuran en el programa oficial, o a elevadas capacidades intelectuales que no se ha procurado despertar en forma paciente y sistemática durante el año. ¿Cuántos fracasos escolares no se deben a las lindas preguntas de "inteligencia" planteadas al término de una enseñanza que no se esforzó en cultivarla?

Si bien la definición precisa de los objetivos permite una validación del contenido mucho más adecuada que en el pasado, no por eso deja de ser aún muy limitada en el estado actual de nuestros conocimientos. En primer término, porque es muy difícil tener la certeza de que tal comportamiento real es verdaderamente representativo de tal rasgo que aspirábamos a cultivar. Luego, porque aun si la visión es correcta en un principio, las vicisitudes de la calificación siempre pueden traicionar las intenciones.

Un profesor de historia o de geografía, escribe P. Vernon², que por medio de un examen escrito trata de juzgar la comprensión profunda de esas disciplinas, asigna a menudo

² P. Vernon, "Types of Examination", en J. Lauwerys y D. Scanlon, *Examinations*, pág. 43.

una buena parte de los puntos a la reproducción exacta de hechos detallados, a la de sus teorías favoritas, a la extensión de las respuestas, a la inteligibilidad y elegancia de estilo con que se expresan. En un examen oral, ¿en qué medida no dependen los puntos de la inteligencia social, del autocontrol y, en particular, de la aptitud del alumno para crear, mediante la palabra y la actitud, una buena relación con el examinador?

¿Cómo obrar en la práctica?

En primer término, exigir que el profesor indique, en una ficha especial, el objetivo perseguido con cada pregunta de examen que redacta.

Después, someter las preguntas y el esquema de calificación a varios educadores experimentados (uno de los cuales, por lo menos, ha de pertenecer a un grupo de disciplinas que no formen parte del área en cuestión) que juzgarán, en forma independiente, la validez del contenido de las preguntas. En caso de desacuerdo grave, la pregunta será rechazada³.

La experiencia enseña que actualmente hay que atenerse a un reducido número de objetivos generales, concebidos de una manera bastante amplia. En caso contrario, los jueces no logran ponerse de acuerdo. Ya hemos expresado anteriormente que nuestros conocimientos en materia de validación son todavía rudimentarios.

En ciertos casos, y por la misma razón, no se llegará a acuerdo alguno acerca de las cuestiones fundamentales. Sin duda, esto no es motivo suficiente para suspender la enseñanza de la materia o de la asignatura incriminada, pero entonces debe indicarse con toda claridad la imposibilidad de llegar a un acuerdo.

Se corre el riesgo, por ejemplo, de encontrar semejante dificultad en el control sistemático de la realización de los objetivos asignados a las lenguas muertas, a ciertas ramas de las matemáticas, etcétera.

Una vez más comprobamos que al tratar de acrecentar el rigor científico de los exámenes, las enseñanzas serán enjuiciadas. Solo a ese precio desaparecerá el fetichismo pedagógico.

³ Un coeficiente de concordancia entre las opiniones puede calcularse mediante la fórmula de Kendall y, de común acuerdo, elegirse un umbral de aceptación.

II. La validez de predicción

Resulta, aparentemente, la más sencilla de controlar, porque no exige ninguna comprensión profunda de los fenómenos. Aun si no nos explicamos la relación entre el éxito en un determinado *test* y un brillante desempeño en cierto terreno, el hecho es fácil de observar.

Se trate de un examen de admisión en un nuevo ciclo de estudios o de autorización para el ejercicio de una profesión, basta con seguir a los individuos durante cierto tiempo para saber si el pronóstico formulado a partir de las notas otorgadas se verifica o no.

Comprobamos con asombro que tal verificación pocas veces se hace. J. C. Passeron advierte en esto la señal de que el examen no se destina tanto a medir objetivamente la capacidad como a satisfacer las necesidades *sociales* de las clases privilegiadas⁴. Tal afirmación ha de aceptarse con reservas, por cierto, pero hemos visto, en la primera parte, que no carece enteramente de fundamento.

Para que no sea un engaño, la verificación sistemática de la validez predictiva exige el control riguroso de las variables; en caso contrario, las variables ocultas falsificarán las conclusiones. Por ejemplo, hasta hace poco se argumentaba todavía a favor del latín y se apelaba para ello a la estadística del número de estudiantes sobresalientes que habían cursado los estudios secundarios clásicos. Ahora bien, lo único que esa observación demuestra es que en nuestros países, los alumnos más inteligentes son orientados, por tradición, y a partir de los doce años, hacia ese tipo de estudios.

A manera de síntesis, tomamos la siguiente recapitulación del *Examination Bulletin* n° 3⁵.

FACTORES DE VALIDEZ DE UN EXAMEN

1. Identificación adecuada de los objetivos cuya realización ha de verificarse mediante el examen escrito, la evaluación de la tarea del año, el trabajo práctico, etcétera.
-

⁴ J. C. Passeron, *Sociologie des examens*, op. cit., pág. 7.

⁵ Londres, H. M. S. O., 1964, págs. 19-20.

-
2. Entre toda la gama de objetivos identificados de esta manera, elegir los que han de concentrar la verificación.

 3. Evaluación eficiente de la adecuación del contenido y de la estructura del examen al objetivo perseguido.

 4. Relación clara entre cada pregunta y los objetivos de la enseñanza.

 5. Elaboración de un esquema de calificación y redacción de directivas para los encargados de corregir, en función de los objetivos. Además, debe asegurarse un nivel de fidelidad suficiente.

 6. Un buen conocimiento de las capacidades de los candidatos.

 7. Buena disposición de la gente que califica para tener en cuenta juicios independientes de los suyos y datos objetivos que les fuesen proporcionados.

 8. Comparación con pruebas anteriores cuya validez está demostrada.

CUARTA PARTE

LOS PROCEDIMIENTOS DE MODERACION

PLANTEO DEL PROBLEMA

1. Definición

En docimología, la *moderación* ha tenido primeramente la finalidad de mitigar los excesos de severidad o generosidad de algunos examinadores.

En un sentido amplio, designa en la actualidad al conjunto de medidas tomadas para hacer comparables las notas de exámenes internos y, por consiguiente, para unificar su significado a nivel de las distintas divisiones del mismo tipo en una escuela, en un grupo de escuelas, en establecimientos similares de una región o de un país.

Las investigaciones más sutiles acerca de la moderación son obra, ciertamente, de los docimólogos británicos. En un sistema completamente descentralizado como el de Gran Bretaña, los programas y métodos de las escuelas primarias y secundarias pueden diferir considerablemente. Por esta razón, durante muchísimo tiempo, exámenes externos, es decir, organizados fuera de las escuelas por comisiones especiales, han tenido que sancionar los estudios en momentos cruciales de la escolaridad. Es conocido el famoso 11 + *Examination*, examen de ingreso en el colegio secundario que decide, en particular, sobre la admisión en la enseñanza general clásica, y el C. S. E. (*Certificate of Secondary Education*), antes de entrar en la enseñanza superior. Este último examen, por ejemplo, provee a cada alumno de un documento que indica el nivel alcanzado en cada una de las materias, de acuerdo con normas nacionales.

Para admitir a un estudiante en una determinada sección, las universidades anuncian sus exigencias especiales: una

requiere notas sobresalientes (A) en matemáticas, biología y lengua materna; otra, acepta igualmente la nota "muy bien" (B); etcétera.

Ese sistema de exámenes externos se critica, a veces muy severamente, porque limita en gran medida la libertad pedagógica de los maestros y de las autoridades locales.

Por esta razón, desde hace algunos años, se trata de sustituirlos por exámenes internos cuyos resultados se hacen comparables mediante procedimientos de moderación local, regional, y luego nacional, elaborados en forma experimental.

En países como Bélgica, donde un programa único de estudios es impuesto por una autoridad central, resulta mucho más fácil moderar los exámenes, si esto es lo que se desea. Antes de señalar en qué forma, tal vez no sea superfluo recordar por qué es deseable una moderación.

En términos generales, existe una doble razón. La primera concierne a los individuos, alumnos y padres, que tienen el derecho de conocer el nivel "real" de los rendimientos escolares, antes de decidir acerca de la orientación ulterior en los dominios de los estudios o de la profesión. ¿Quién se consideraría calificado para los juegos olímpicos por haber ganado una carrera organizada entre algunos amigos? Por lo demás, una comprobación desfavorable no representa una condena sin apelación, no es más que un diagnóstico de debilidad o enfermedad. Hay que saber a tiempo que se plantea un problema para buscar sus causas y luego los remedios, si existen.

La segunda razón se refiere a la comunidad. Así como un consumidor no ha de ser engañado con respecto a la mercadería, tampoco la sociedad puede ser obligada a pagar muy caros los fracasos de un estudiante incorporado a la universidad sobre el testimonio de un certificado inválido, ni a confiar una función a alguien que no la merece. El día en que, por ejemplo, los exámenes de la enseñanza normal sean rigurosamente controlados, cabe esperar que ya no se confíen generaciones de niños a ciertos maestros cuyos conocimientos de la lengua materna, sobre todo de la ortografía¹, son más que insuficientes.

¹ No padecemos del fetichismo de la ortografía, pero nos parece imposible que la desconozca quien debe enseñarla.

2. Moderar no es someter a disciplina rígida

Podría temerse que el deseo de hacer comparables los resultados escolares provocara un resurgimiento de las antiguas restricciones. Recordamos a aquel inspector de tradición napoleónica que, consultando su reloj, creía hallarse en condiciones de afirmar: "En este instante se enseña tal lección en todos los quintos grados primarios de Francia". Conocemos los defectos de una enseñanza militarizada.

Pero la necesidad de cierta plasticidad pedagógica no impide que, más allá de las adaptaciones circunstanciales, cada ciclo de estudios persiga objetivos fundamentales, comunes a todos: la adquisición de técnicas básicas o de conocimientos y capacidades que se consideran esenciales. Maestros y moderadores tienen que interrogarse y ponerse de acuerdo sobre esos aprendizajes esenciales.

3. ¿Moderación espontánea o impuesta?

Normalizar la totalidad de los exámenes en todos los momentos de los estudios no parece deseable ni, por otra parte, posible. En el primer caso, porque los alumnos progresan con ritmos a veces muy diversos y se adaptan mal a las divisiones rigurosas en períodos y años escolares. Y resulta imposible porque la moderación, incluso mediante los métodos económicos que describiremos, sigue siendo un trabajo arduo en cuanto a preparación y ejecución. La moderación debe imponerse en los momentos decisivos, sobre todo al otorgar diplomas o certificados de fin de estudios.

Por otra parte, la decisión debería estar en manos de los maestros en forma individual, o bien de los directores de los establecimientos o de las autoridades locales.

4. La moderación empieza junto con el año lectivo

Como acabamos de señalar, es importante que los profesores y moderadores hayan podido ponerse de acuerdo acerca de los objetivos de la enseñanza y de algunos grandes principios de la calificación, a fin de que los exámenes resulten comparables.

Vagas afirmaciones, tales como "El alumno debe ser capaz de expresarse correctamente por escrito", son inútiles. La capacidad deseada ha de traducirse a términos de comportamientos concretos que, por lo demás, serán incorporados en las escalas de evaluación descriptivas. Si se trata de la expresión escrita, se plantean una serie de interrogantes:

¿Cómo graduar las exigencias en función de la edad de los alumnos y del tipo de escuela?

¿Será recompensada la originalidad? ¿Cómo? ¿Cómo se la identificará?

¿Qué importancia se atribuirá a la ortografía?

¿Se tomará en cuenta la extensión de los trabajos de redacción? ¿Se exigirá una extensión mínima, por debajo de la cual se descontarán puntos? ¿Cuántos?

¿Qué importancia se atribuirá a la riqueza del vocabulario?

¿Cómo definirla? ¿Cuáles serán las exigencias con respecto a la sintaxis?

Los matemáticos pueden, asimismo, "afinar sus instrumentos". Si algunos profesores, para limitarnos a un solo ejemplo, castigan gravemente los errores de operación, y hasta las faltas de ortografía, en la solución de los problemas, mientras sus colegas opinan que esos aspectos son de menor importancia, se corre el riesgo de que notas iguales cubran realidades muy diferentes.

Los programas oficiales también pueden desempeñar un papel importante mediante la máxima precisión de los objetivos por alcanzar en cada escalón. La eficacia de esos datos básicos resulta tanto mayor cuanto más centralizado se halle el sistema escolar.

Es inútil alentar ilusiones: los objetivos de comportamiento de las distintas materias no se definen en algunas horas de meditación. Para llegar a un resultado satisfactorio, distintas comisiones integradas por docentes, inspectores e investigadores reunirán sus esfuerzos y trabajarán durante mucho tiempo. Además, muchos interrogantes pueden permanecer, por ahora, sin respuesta.

La definición de los objetivos será objeto de una indagación permanente, no solo porque podrán precisarse en función

de los adelantos de la psicología del aprendizaje, sino también porque los mismos objetivos evolucionan.

Las instrucciones relativas a los exámenes se precisarán en forma paralela. Además de las planillas de objetivos generales y especiales, los programas escolares del futuro determinarán claramente el número y los tipos de preguntas por hacer en los exámenes, y ofrecerán indicaciones en cuanto a las líneas de fuerza de los esquemas de calificación.

5. No existe comparación posible sin elevada fidelidad

No es concebible comparar entre sí resultados de exámenes que, tomados en forma aislada, resultarían sumamente fluctuantes.

El problema de la fidelidad ya se discutió en la parte dedicada a la construcción del examen. No volveremos sobre este punto.

6. ¿Podemos confiar en los "tests"?

En los sistemas de moderación que vamos a examinar, los *tests* de inteligencia o de conocimientos ocupan un lugar importante. En ciertas ocasiones, aportan los puntos de referencia a los cuales se ajustan los demás resultados. En otros casos, se les atribuye tanto valor como al trabajo del año y al examen final.

Como ya hemos visto con qué rigor y precauciones se construyen los buenos *tests* de conocimientos, no ha de causarnos extrañeza la confianza que se otorga a esos instrumentos. Por lo demás, desde la primera parte de este libro hemos visto asimismo que si los maestros juzgan bien a sus alumnos, un simple *test* verbal, administrado en menos de una hora, permite una predicción bastante segura:

	<i>Correlación con los resultados del alumno:</i>	
	<i>2 años después</i>	<i>3 años después</i>
Pronóstico del maestro	0,821	0,748
Test verbal	0,796	0,722

Aquí tenemos algunos resultados de investigaciones que confirman esas observaciones.

Término medio, un *test* de inteligencia o, mejor aún, una combinación de puntajes obtenidos en distintos *tests* de inteligencia resultan mejores "predictores" del éxito escolar que los resultados de los exámenes. Tal superioridad quedó demostrada una y otra vez, sobre todo por investigadores anglosajones:

- 1) Emmet² muestra que un *test* de inteligencia verbal, después de dos o tres años de enseñanza secundaria general, permite prever mejor los resultados que los exámenes de lenguaje y aritmética evaluados por los maestros.
- 2) Más tarde, Emmet y Wilmut³ hicieron una demostración igualmente convincente del valor "predictivo" del *test* de inteligencia cinco años después.
- 3) Wrigley⁴ confirmó esos resultados y mostró que la predicción puede ser aún mejor si existe la posibilidad de combinar los resultados de un *test* de inteligencia y de los *tests* de conocimientos estandarizados.
- 4) En Bélgica, todos los orientadores conocen el valor "predictivo" de los puntajes verbales y de razonamiento del *test* P. M. A. de Thurstone, el famoso 2V + R, para la enseñanza secundaria general.

No disponemos de investigaciones similares para la enseñanza técnica o para estructuras de enseñanza tal vez mejor adaptadas a la civilización del año 2000. Es muy posible que debieran utilizarse los puntajes de otros tipos de *tests*.

² W. Emmet, *An Inquiry into the Prediction of Secondary School Success*, Londres, Univ. of London Press, 1942.

³ W. Emmet y F. Wilmut, "The Prediction of School Certificates Performance in Specific Subjects", en *British Journal of Educ. Psychol.*, 22, 1952, págs. 52-62.

⁴ J. Wrigley, "The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools", en *British Journal of Educ. Psychol.*, 25, 1955.

ALGUNOS SISTEMAS DE MODERACION DE EXAMENES

I. La fórmula más liberal:

El sistema sueco de moderación por materia
a partir de un "test" de conocimientos

Desde el fin de la segunda Guerra Mundial, Suecia adoptó un sistema de moderación, sencillo y optativo, aplicable por cualquier profesor en su clase, a lo largo de toda la escolaridad.

La totalidad de las notas de los profesores se asignan de acuerdo con una escala de 7 grados¹.

Se propone que el porcentaje de alumnos que reciben una determinada nota sea, de una manera general, el que reproducimos a continuación² y que deja al profesor en libertad de tomar en cuenta características de su clase (por ejemplo: un alumno excelente a la cabeza o casi todos los alumnos regulares, etcétera):

Notas	1	2	3	4	5	6	7
Porcentaje	1	6	24	38	24	6	1

¹ Nos referimos a la exposición muy clara de S. Henrysson, "The Swedish System of Equalising Marks", en *Educational Research*, VI, 2, feb. 1964, págs. 156-160.

² Distribución normal (véase curva de Gauss).

Además, todos los años, una oficina central donde colaboran especialistas y profesores experimentados construye una batería de *tests* de conocimientos referentes a los principales puntos del programa. Esa batería se estandariza de acuerdo con una muestra nacional representativa; los resultados se distribuyen en 7 clases, en la proporción aconsejada por las notas de los profesores.

Si el profesor administra el *test* nacional en su clase, dispone, pues, de normas directamente comparables con las notas que ha asignado. Por lo tanto, le resulta fácil ajustar estas últimas.

He aquí un ejemplo concreto propuesto por S. Henrysson:

24 alumnos			
Notas	<i>Distribución de las notas preliminares asignadas por el profesor</i>	<i>Distribución de las notas obtenidas en el test</i>	<i>Distribución de las notas después del ajuste</i>
7	—	1	—
6	—	4	4
5	4	4	4
4	4	9	10
3	14	5	4
2	2	1	2
1	—	—	—
<i>Media</i>	3,42	4,33	4,17

El profesor anota, en primer término, en la 2a. columna, las notas que ha asignado. La media (3,42) ofrece una primera indicación, si la media nacional es conocida. (En Suecia es de 4 para el 6º grado primario. Según parece, en este caso el profesor es demasiado severo.)

Aquí tenemos los resultados obtenidos en el *test* por los mismos alumnos:

<i>Puntaje bruto</i>	<i>Nota correspondiente</i>	<i>Número de alumnos que obtuvieron esa nota</i>
94-100	7	1
84-93	6	4
67-83	5	4
46-66	4	9
30-45	3	5
22-29	2	1
0-21	1	—

Estos resultados se anotan en el primer cuadro, 3a. columna.

La media es de 4,33, y esto confirma la primera impresión de severidad.

Sin cambiar el orden de calificación inicial, el profesor rectifica sus notas (columna 4).

El sistema sueco atrae por varias razones:

1. Los maestros conservan toda su libertad:
 - a) de administrar o no el *test* (casi todos lo hacen);
 - b) de tener en cuenta los resultados;
 - c) de comunicar los resultados de los *tests* a los alumnos, a los otros profesores, al director, a los padres.

De esta manera, el *test* se convierte simplemente en un instrumento puesto a disposición de quienes lo utilizan. Se reduce entonces al mínimo la tentación de preparar los exámenes con el único propósito de salir del paso.

2. Los constructores de *tests* procuran abarcar una amplia gama de objetivos con un buen muestreo de todo el programa.

Es evidente que los profesores prestan una particular atención a las materias tratadas en los *tests*. En consecuencia, se dispone de un medio eficaz para sensibilizar al maestro frente a ciertas innovaciones.

Cabe señalar que, desde 1965, un sistema muy similar se

halla en experimentación en Ginebra, en 5° y 6° grado primario (10-12 años), para ortografía y aritmética³.

Entre las críticas formuladas contra ese sistema de ajuste⁴, cabe mencionar:

1. Si un profesor ha juzgado a uno de los alumnos de su clase con excesiva severidad, la referencia a la prueba nacional no reparará la injusticia.
2. La prueba de referencia tal vez no mida las mismas dimensiones que las notas.
3. La construcción de la prueba de referencia puede ser, asimismo, deficiente: escasa fidelidad, . . .

**II. Sistema de moderación establecido
a partir de un "test" de conocimientos
y de acuerdo con los puntos principales
del programa**

A partir de los mismos datos básicos que los de Suecia (exámenes internos y *tests* de conocimientos para las materias principales, estandarizados a nivel nacional o regional), el siguiente sistema directivo puede brindar buenos resultados.

Una comisión nacional o regional de moderación se hace cargo de cierto número de escuelas.

Estas le remiten los resultados obtenidos en las pruebas preparadas libremente por los profesores, así como los puntajes de los *tests* de conocimientos.

En caso de haber diferencia, en más o en menos, igual o superior a dos desvíos típó (por ejemplo) entre las medias de los exámenes y de los *tests*, la escuela recibe la visita de los moderadores.

Su misión no consiste en decir al director cómo debe conducir su escuela, ni al profesor cómo llevar su curso, sino en llamar la atención sobre un hecho y tratar de hallar la

³ Véase S. Roller, "Le problème de l'attribution des notes scolaires. Essai de solution", Ginebra, en *Docimologie et éducation*, número especial de la revista *Les sciences de l'éducation*, abr.-set. 1969, pág. 66 y sigs.

⁴ W. Angoff, "Can Usual General Purpose Equivalency Tables be Prepared for Different College Admission Tests"; en A. Anastasi, ed. *Testing Problem in Perspective*, Washington, ACE, 1966, págs. 251-264. F. Bacher, *La normalisation des notes*, *op. cit.*, pág. 63.

explicación y, si es posible, remedio, en plena colaboración con la escuela.

Si el desacuerdo subsiste, cabe la posibilidad de otorgar a las autoridades regionales el derecho de ajustar las notas oficialmente, y a la escuela, el de interponer recurso.

III. Moderación espontánea mediante apelación a un "banco" de preguntas

Cuando los maestros recurren a preguntas en forma de ítem de elección múltiple, parece muy promisorio el siguiente procedimiento, que actualmente se experimenta en Gran Bretaña⁵. Este requiere, sin embargo, la existencia de un servicio de investigación, necesidad obvia, por otra parte, para cualquier experto educador de nuestros días.

1. Los maestros indican en una planilla los objetivos perseguidos.
2. Remiten la planilla al servicio de investigación junto con los ítem que han redactado y, quizá, ya "testeados" en forma local.
3. El servicio examina esos ítem y, según las posibilidades, pone a prueba algunos de ellos. Se calculan su dificultad y poder discriminativo.
4. El servicio devuelve al maestro:
 - a) los ítem examinados;
 - b) algunos ítem complementarios cuyos índices de facilidad y eficacia se conocen con respecto a determinada población: ciudad, cantón, país, . . .Esos ítem ofrecerán un punto de comparación en función del cual será factible ajustar el resultado registrado para el resto.

Es interesante señalar que, después de haberse parapetado fuertemente en la redacción de ítem de respuesta cerrada, mediante elección múltiple, los británicos empiezan a proponer preguntas semiabiertas. De hecho, se trata de preguntas abiertas, muy cuidadosamente formuladas, para las cuales se identificaron en forma experimental los grandes tipos de

⁵ Véase D. Pidgeon y A. Yates, *op. cit.*

respuestas, en función de las cuales se proponen tablas de corrección (baremos).

La organización de un banco de ítem es ardua tarea durante los primeros años. Después, el caudal de ítem estandarizados alcanza tal magnitud que el trabajo se alivia y el sistema adquiere gran flexibilidad y rapidez de funcionamiento.

IV. Un sistema sencillo de moderación global al servicio de la selección

Ya observamos que, por lo general, los maestros juzgan en forma correcta a sus alumnos, pero tienden a relativizar sus juicios en relación con el nivel global de la clase.

El siguiente sistema, elaborado por la Fundación Nacional Inglesa para la Investigación en Educación⁶, destinado a la admisión en la enseñanza secundaria general clásica, permite *seleccionar* teniendo en cuenta juicios independientes de los maestros:

1. El maestro clasifica a sus alumnos por orden de mérito, es decir, según su evaluación global de la posibilidad de éxito en la enseñanza secundaria general. Varios alumnos pueden clasificarse *ex aequo** (clasificación 1).
2. Los alumnos se someten a un *test* verbal de inteligencia (2).
3. Los puntajes del *test* se clasifican por orden decreciente (clasificación 3).
4. Las clasificaciones (1) y (3) se yuxtaponen.
5. El puntaje de inteligencia que se encuentra frente al nombre del alumno es considerado un puntaje de evaluación por el maestro (juicio ajustado).

⁶ Véase A. Yates y D. Pidgeon, *Admission to Grammar Schools*, *op. cit.*

**Ex aequo*: con igualdad, en condiciones iguales (*N. del T.*)

<i>Clasificación por el maestro (1)</i>	<i>Puntaje obtenido por el alumno en el test verbal (2)</i>	<i>Clasificación de los puntajes por orden decreciente (3)</i>	<i>Juicio ajustado (4)</i>
A	121	132	132
B	120	128	128
C	132	121	121
D	128	120	120
E	100	106	106
F	106	100	100
G	94	100	97
H	82	96	97
I	96	94	97
J	100	86	86
K	78	82	78
L	79	79	78
M	86	78	78
N	73	73	78
O	65	65	65
		Total	290
		Total	312

Vemos que la clasificación a la cual se llega finalmente no modifica de manera alguna el orden elegido en un principio por el maestro. Pero esta vez disponemos de un medio de comparación entre las escuelas.

Utilización de juicios ajustados

Ejemplo: En un colegio secundario se desea reclutar, para el 6° año de latín, a 25 alumnos provenientes de tres escuelas primarias distintas.

Juicios ajustados

	Escuela primaria I	Escuela primaria II	Escuela primaria III
	132		
130			130 x 2
	128		126
125		125 124	125
	121	121	123
120	120	119	120
		117	116
115	115 x 3	115	
	112		114 113
110		110	110
		107	
105	106	104	
100	100		

Para seleccionar a 25 alumnos, cortamos, en este caso, en el nivel 110. Si en ese nivel de 110, los *ex aequo* hubiesen ubicado a más de 25 alumnos en la categoría de los seleccionados, la más elemental justicia habría exigido admitirlos a todos.

Trazar una línea de demarcación como acabamos de hacerlo representa una fuente de injusticia por otra razón: ¿no se colocaría cierto número de sujetos justo por encima o por debajo del límite, por simple azar?

El psicómetro está preparado para salvar esa dificultad: toma un margen de seguridad equivalente a tres veces el error

estándar que una sencilla fórmula permite calcular en función del coeficiente de fidelidad del *test* utilizado⁷.

V. En Inglaterra, un sistema completo de moderación

El sistema que vamos a describir ahora proviene, asimismo, de Gran Bretaña. Por lo que sabemos, no existe otro más completo: abarca la preparación del examen en colaboración con las escuelas, la calificación, y el ajuste de las notas y de los grados finales.

Su objetivo estriba en *perfeccionar los exámenes internos*, a tal punto, que se les pueda confiar el papel desempeñado hasta por las grandes pruebas externas.

Para llegar a una moderación a nivel nacional se procede por etapas: a nivel local se agrupa cierto número de escuelas que persiguen un mismo fin; luego, a nivel regional, de acuerdo con una técnica que no difiere fundamentalmente de la aplicada en el ámbito local; para el paso a nivel nacional se agrega, a la técnica de moderación propiamente dicha, un muestreo detallado cuyo estudio técnico no viene al caso hacer en este lugar.

En cuanto a nuestro país, proponemos que algunas escuelas animadas por el mismo espíritu empiecen por hacer un ensayo voluntario con el sistema. Por esa razón concentramos nuestra presentación sobre el proceso de base.

La moderación interescolar solo se haría en los momentos cruciales de la escolaridad y respecto de algunas materias principales. Sin embargo, nada impide que los profesores de una misma escuela, que tienen a su cargo los mismos cursos, utilicen espontáneamente idéntico método de corrección.

A. Preliminares

Ya mencionamos la necesidad de que exista acuerdo en cuanto a los objetivos y la construcción del examen. No volveremos sobre este punto.

⁷ Error estándar = $\sigma \sqrt{1-r}$

siendo σ = desvío tipo de los puntajes.

r = coeficiente de fidelidad. Este coeficiente suele indicarse en el manual que acompaña el *test*.

Cada escuela elige entre su cuerpo docente a un profesor que desempeñará dos papeles:

1. Coordinador de los exámenes en su escuela.
2. Miembro de la comisión de moderación interescolar.
No existe ninguna imagen ideal del moderador. Las siguientes cualidades parecen deseables⁸:
1. Madurez general y buena experiencia pedagógica.
2. Frecuentes contactos con alumnos del nivel y del tipo de escuela correspondientes al examen.
3. Capacidad para formular con claridad sus criterios y juicios.
4. Bondad con firmeza.
5. Capacidad para discutir sin apasionamiento, sin crear tensiones.
6. Aptitud para comprender algunas técnicas de análisis estadístico.

Las preguntas de exámenes se redactan en cada escuela con gran libertad de movimiento: solo la lista de objetivos principales y el número y la extensión de las preguntas deben respetarse según lo convenido.

Por lo menos en un principio, conviene someter las preguntas a la comisión de moderación, que se cerciorará de la unidad general.

B. Los profesores califican sus exámenes

C. Nueva corrección de una muestra de trabajos por parte de los moderadores

Tres principios rigen el trabajo:

1. La intervención en los exámenes de las escuelas ha de resultar tan discreta como sea posible. Solo los casos de divergencias importantes merecen atención.
2. Los encargados de la corrección deben ser, de una manera general, de igual severidad, apreciar las mismas cualidades y

⁸ Mather, *op. cit.*, pág. 67.

estar de acuerdo respecto del significado de los grados finales.

3. Las muestras que se volverán a corregir han de ser pequeñas y el trabajo estadístico resultará todo lo sencillo que sea posible.

En cada escuela, y para un mismo tipo de examen, se extraen al azar 20 trabajos corregidos.

Supongamos que se trate de 12 liceos⁹. Cada uno ha delegado un moderador en la comisión, y ésta recibe, en consecuencia, 12 paquetes de 20 trabajos.

La primera operación tiende a determinar la equivalencia de los moderadores desde tres puntos de vista:

1. *Severidad*: el grado de severidad se revela por la media.
2. *Discriminación*: calificación demasiado libre o excesivamente prudente. Revelada por la dispersión o el margen de variación de las notas.
3. *Conformidad*: ¿todos los encargados de la corrección califican de la misma manera al mismo alumno? Revelada por la correlación entre dos series de notas.

Para verificar la coincidencia entre moderadores desde estos tres puntos de vista, los doce vuelven a corregir primero un mismo paquete de 20 trabajos.

Los 20 trabajos se distribuyen, siempre al azar, en cinco grupos de cuatro. En cada grupo, los alumnos son calificados por orden alfabético.

Luego se realizan las sencillas operaciones que vienen a continuación, en el orden en que figuran en el cuadro de la página siguiente.

⁹ Ejemplo tomado del *Examinations Bulletin* n° 5, Londres, HMSO, 1965.

ESTUDIO DEL ACUERDO ENTRE LOS MODERADORES

Notas otorgadas por 12 moderadores

Nombre del candidato	MODERADORES												Media
	a	b	c	d	e	f	g	h	i	j	k	l	
A	4	4	4	3	4	4	3	1	4	4	4	4	4
B	2	2	2	2	3	2	1	1	2	2	2	2	2
C	6	5	5	5	5	5	5	5	5	5	5	5	5
D	2	1	1	2	2	3	3	2	2	3	2	3	2
Total	14	12	12	12	14	14	12	9	13	14	13	14	13
Margen de variac.	4	4	4	3	3	3	4	4	3	3	3	3	3
E	3	3	3	4	3	5	3	3	3	3	3	3	3
F	1	1	1	1	1	1	1	1	1	1	1	1	1
G	5	6	6	6	5	3	6	6	6	6	6	6	6
H	3	4	4	4	3	4	3	4	4	2	4	4	4
Total	12	14	14	15	12	15	13	14	14	12	14	14	14
Margen de variac.	4	5	5	5	4	4	5	5	5	5	5	5	5
I	2	2	2	2	4	3	2	3	3	3	3	3	3
J	2	1	2	1	3	2	2	2	2	2	2	2	2
K	5	5	5	6	5	5	5	6	5	5	5	4	5
L	1	2	2	1	2	2	2	3	2	2	3	2	2
Total	10	10	11	10	14	12	11	14	12	12	13	11	12
Margen de variac.	4	4	3	5	3	3	3	4	3	3	3	2	3
M	4	2	3	4	3	4	5	5	5	4	5	4	4
N	4	3	4	4	4	4	5	4	4	4	4	4	4
O	5	4	4	3	4	3	4	4	4	5	4	5	4
P	5	6	5	5	5	5	5	5	5	5	5	5	5
Total	18	15	16	16	16	17	18	18	18	18	18	18	17
Margen de variac.	1	4	2	2	2	2	1	1	1	1	1	1	1
Q	1	2	1	1	1	1	1	1	1	1	2	1	1
R	1	1	2	1	1	2	1	1	1	1	1	1	1
S	3	3	2	3	3	2	2	3	3	3	3	3	3
T	2	1	2	1	2	2	1	2	2	2	2	2	2
Total	7	7	7	6	7	7	5	7	7	7	8	7	7
Margen de variac.	2	2	1	2	2	1	1	2	2	2	2	2	2
Total general	61 ²	58	60	59	63	65	59	62	64	63	66	64	63 ¹
Suma de márgenes	15 ⁴	19	15	17	14	13	14	16	14	14	14	13	14 ³

Diferencias entre las notas otorgadas por cada moderador y la media de las notas dadas por la totalidad de los moderadores a un mismo candidato

Nombre del candidato	MODERADORES											
	a	b	c	d	e	f	g	h	i	j	k	l
A	0	0	0	-1	0	0	-1	-3	0	0	0	0
B	0	0	0	0	1	0	-1	-1	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0	0	0
D	0	-1	-1	0	0	1	1	0	0	1	0	1
Margen de variac.	1	1	1	1	1	1	2	3	0	1	0	1
E	0	0	0	1	0	2	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0
G	-1	0	0	0	-1	-1	0	0	0	0	0	0
H	-1	0	0	0	-1	0	-1	0	0	-2	0	0
Margen de variac.	1	0	0	1	1	3	1	0	0	2	0	0
I	-1	-1	-1	-1	1	0	-1	0	0	0	0	0
J	0	-1	0	-1	1	0	0	0	0	0	0	0
K	0	0	0	1	0	0	0	1	0	0	0	-1
L	-1	0	0	-1	0	0	0	1	0	0	1	0
Margen de variac.	1	1	1	2	1	0	1	1	0	0	1	1
M	0	-2	-1	0	-1	0	1	1	1	0	1	0
N	0	-1	0	0	0	1	0	0	0	0	0	0
O	1	0	0	-1	0	-1	0	0	0	1	0	1
P	0	1	0	0	0	0	0	0	0	0	0	0
Margen de variac.	1	3	1	1	1	2	1	1	1	1	1	1
Q	0	1	0	0	0	0	0	0	0	0	1	0
R	0	0	1	0	0	1	0	0	0	0	0	0
S	0	0	-1	0	0	-1	-1	0	0	0	0	0
T	0	-1	0	-1	0	0	-1	0	0	0	0	0
Margen de variac.	0	2	2	1	0	2	1	0	0	0	1	0
Suma de márgenes ⁵	4	7	5	6	4	8	6	5	1	4	3	3

Controles

Las siguientes reglas no tienen nada de mágico. Para formularlas, G. Peaker se inspiró en las técnicas de control de calidad empleadas en la industria. Experimentos muy avanzados en el terreno de los exámenes han permitido su adaptación.

El control realizado de esta manera resulta razonable, expeditivo, eficaz, pero no muy preciso. Por esta razón, si a pesar de la amplitud de criterios, un aspecto de la corrección parece inaceptable, existe sin duda un problema.

1. Severidad

La experiencia muestra que los esfuerzos de armonización deben orientarse, ante todo, hacia ese punto. Según veremos, en caso de un problema, la solución es fácil, felizmente.

Regla:

- a) Partir de la media¹⁰ de los totales generales (en este caso, 63).
- b) No tolerar desvíos más allá de una variación central de 10 puntos¹⁰ con respecto a esa media, es decir, 5 puntos en más o en menos (aquí: 58-68).

Comprobación:

Todos los totales generales están comprendidos dentro de esa variación.

Todos los que corrigen son, por lo tanto, de una severidad aceptable.

Solución en caso de exceso de severidad o de generosidad. Supongamos que nos vemos frente a los siguientes totales: 75-72-68-66-65-64-60-49-46-44-41-40.

Comprobamos que si se reúne al más severo con el más generoso, luego al segundo en severidad con el segundo en generosidad, y así sucesivamente, las medias de los pares se hallan muy cerca de la media general.

<i>Sea:</i>			<i>Media</i>
(1)	40	75	57,5
(2)	41	72	56,5
(3)	44	68	56
(4)	46	66	56
(5)	49	65	57
(6)	60	64	62
Media general			57,5
Variación de 10 puntos			52,5-62,5

Únicamente los dos encargados de la corrección del par n° 6 se sitúan entre 52,5 y 62,5. Ellos serán autorizados a

¹⁰ Los encargados de corregir califican 20 redacciones de 1 a 5. Diez puntos equivalen a una diferencia media de medio punto por trabajo para el conjunto de los 20 trabajos.

corregir solos. En los otros casos, los trabajos serán sometidos a los dos miembros de uno de los grupos constituidos y se calculará la media de los puntos.

2. *Discriminación*

Regla:

El total de las variaciones medias (3) no puede ser superior al doble del total de las variaciones del que corrige (4), y viceversa.

Comprobación:

No hay problema.

3. *Conformidad*

Regla:

El total de variaciones de las diferencias (5) entre las notas asignadas por un moderador y la media de las notas asignadas por el conjunto no ha de ser superior a 12¹¹.

Comprobación:

No hay problema.

Conclusión

En el presente caso, todos los moderadores se mantuvieron dentro de los límites de las tres pruebas de control. Pueden, por lo tanto, trabajar solos.

¿En qué punto estamos? De las doce muestras de 20 copias, una está corregida ahora definitivamente (¡la han visto doce personas!).

Como cada moderador puede trabajar solo, el resto de las operaciones marchará con rapidez.

D. Nueva corrección de las muestras restantes y control

Las operaciones son prácticamente las mismas que cuando se trata del control de los moderadores. No obstante, ofrecemos un ejemplo detallado porque la presentación más concisa (un solo moderador y un único profesor) proporciona una mejor visión de conjunto.

¹¹ Mostramos en la pág. 177 que con esa marca empírica se obtiene una buena estimación de la correlación entre dos series de notas.

**CALCULOS PARA LA COMPARACION
ENTRE UN PROFESOR Y UN MODERADOR ***
(1 es la nota más alta; 5 la más baja)

5 grupos de 4	Alumnos elegidos al azar	Nota asignada por				Diferencia (Moderador-Profesor)	
		Moderador		Profesor			
1er grupo	Enrique	Min.	5	5	Min.	0	
	Juan		4	4		0	Min.
	Pablo	Max.	2	1	Max.	1	
	Pedro		3	2		1	Max.
	Margen: Min-Max.	3	14	12	4	Margen de las diferencias = 1	
2º grupo	Andrés		2	1	Max.	1	Max.
	Eduardo	Max.	1	1		0	Min.
	Julio		1	1	Min.	0	
	René	Min.	4	4		0	
	Margen: Min-Max	3	8	7	3	Margen de las diferencias = 1	
3er grupo	Antonio		3	3	Max.	0	Max.
	Camilo		5	5		0	
	Eugenio	Max.	2	3		-1 ¹	Min.
	Jerónimo	Min.	6	6	Min.	0	
	Margen: Min-Max	4	16	17	3	Margen de las diferencias = 1	
4º grupo	Santiago	Min.	5	5	Min.	0	
	Lorenzo	Max.	1	1	Max.	0	Min.
	Martín		4	3		1	Max.
	Víctor		4	3		1	
	Margen: Min-Max	4	14	12	4	Margen de las diferencias = 1	
5º grupo	Bruno	Max.	3	2	Max.	1	
	Hugo	Min.	4	2		2	Max.
	León		4	5	Min.	-1	Min.
	Simón		3	2		1	
	Margen: Min-Max	1	14	11	3	Margen de las diferencias = 3	
Los cinco grupos juntos	Total de grados		66	59		Total de márgenes de diferencias = 7	
	Márgenes reunidos	15	(1) (3)	(2) (4)	17	(5)	

* Se considera que (0) es mayor que (-1).

Controles**1. Severidad de los encargados de la corrección**

La diferencia entre total (1) y total (2) no debe ser superior a 10.

Aquí, $66-59 = 7$: aceptable.

2. Discriminación

El total (3) no puede ser superior al doble del total (4), y viceversa.

Aquí, 15 y 17: aceptable.

3. Conformidad

El total de las variaciones de las diferencias (5) no debe ser superior a 12.

Aquí, $= 7$: aceptable¹².

4. Conclusión

En este caso, parece que no se plantea problema alguno. Las notas del profesor son, pues, aceptadas y se terminan las operaciones.

Si las notas del profesor no resultan aceptables, lo ideal es entrar en una discusión con él para tratar de llegar a un acuerdo.

Esto no siempre es posible, sobre todo por falta de tiempo.

La solución siguiente es más expeditiva. Si un moderador encuentra un problema, invita a cuatro de sus colegas a volver a corregir en forma independiente la misma muestra de veinte trabajos. Se calcula la media entre los cinco moderadores y se invita al profesor a ajustar todas sus notas en función de las de los moderadores.

Según G. Peaker, los casos que requieren semejante

¹² Esta manera sencilla de proceder evita el largo cálculo de la correlación sobre las dos series de notas.

Se verá, además, que se obtiene un coeficiente de correlación muy cercano al coeficiente calculado por el método clásico, mediante la operación:

$$\frac{152 + 172 - 72}{2 \times 15 \times 17} = 0,90.$$

corrección son cada vez más raros en la medida en que funciona el sistema de moderación, a causa de los fructuosos intercambios de opiniones que trae aparejados.

E. ¿Cómo ajustar las notas discordantes?

Se trata de un simple ajuste de la media (o de la mediana) por obra de la elevación o descenso de las notas de todos los alumnos (severidad o generosidad excesivas), o bien, asimismo, de una modificación de la distribución general con el propósito de mejorar la discriminación.

1. Ajuste de la mediana¹³

Para un mismo examen, la situación es la siguiente:

	<i>Cuartil superior</i>	<i>Mediana</i>	<i>Cuartil inferior</i>	<i>Desvío tipo aproximado</i>
Moderador	16	13	11	$\sigma = 4$
Profesor	11	8	6	$\sigma = 4$

El profesor es más severo que el moderador, pero discrimina igualmente bien. Si el profesor agrega 5 puntos a cada alumno, el paralelismo se restablece.

2. Ajuste de la mediana y del desvío tipo

Situación inicial:

	<i>Cuartil superior</i>	<i>Mediana</i>	<i>Cuartil inferior</i>	<i>Desvío tipo aproximado</i>
Moderador	16	13	11	$\sigma = 4$
Profesor	16	12	9	$\sigma = 5$

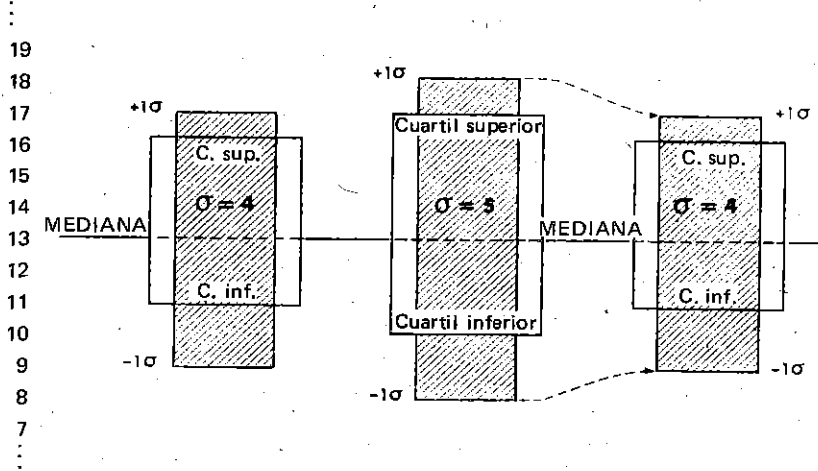
Para igualar con el moderador, se corrige primero la mediana; se agrega un punto a todos los alumnos.

¹³ Adaptado según D. Mather y colab., *op. cit.*, pág. 183.

Supongamos que las notas iniciales del profesor son:
 ... 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 ...

12
 mediana

En primer término mostraremos, mediante el dibujo, en qué consiste el ajuste del desvío tipo.



I. Notas del moderador.

II. Notas del profesor con un aumento de un punto. La mediana llegó a ser igual a la del moderador. Queda por ajustar σ .

III. Notas del profesor disminuidas por segunda vez para alinear el desvío tipo con el del moderador.

Para la operación III, el ajuste en función del nuevo desvío tipo es un poco más complicado que para la mediana. La elaboración de una tabla de conversión facilitará las operaciones.

Para hallar la nueva nota correspondiente a $+1\sigma$, basta con agregar el nuevo sigma a la nueva mediana, o sea, $13 + 4 = 17$. La nueva nota correspondiente a $-1\sigma = 13 - 4 = 9$, etcétera.

Para las notas que corresponden a fracciones de sigma, se calcula primero el desvío en relación con la mediana, se multiplica por $4/5$ y se agrega el resultado a la mediana. El siguiente ejemplo aclara esa fase.

	<i>Notas iniciales del profesor</i>	<i>Notas ajustadas en función de la nueva mediana</i>	<i>Segundo ajuste en función del nuevo desvío tipo (σ)</i>
	.	.	.
	19	.	.
	18	.	.
+ 1 σ	17	18	17*
	16	17	16**
	15	16	15***
	14	15	15****
	13	14	14
Mediana	12	13	13
	11	12	12
	10	11	11
	9	10	11
	8	9	10
- 1 σ	7	8	9
	6	.	.
	5	.	.
	.	.	.
	.	.	.

¿Cómo hemos encontrados estos números?

* Nueva mediana: 13

$$13 + 4 = 17$$

Nuevo desvío tipo: 4

** En la 2a. columna, 17 se halla a 4 puntos de la mediana.

Multiplicar 4 por $4/5 = 3,2$

Mediana + $3,2 = 16,2$ redondeado a 16

*** $16 - 13 = 3$; $3 \times 4/5 = 2,4$; $13 + 2,4 = 15,4$
redondeado a 15

**** $15 - 13 = 2$; $2 \times 4/5 = 1,6$; $13 + 1,6 = 14,6$
redondeado a 15

VI. LA NOTA DE FIN DE AÑO

Trabajo del año + trabajos prácticos + "test"

*Problema*¹⁴

Se desea clasificar a los alumnos en cinco grupos de mérito en el momento de finalizar sus estudios. *A* será el grado superior y valdrá 1, *E* será el grado inferior y valdrá 5.

Se desea tener en cuenta tres elementos, evaluados cada uno según las mismas escalas de cinco grados:

- trabajo del año = T.A.
- trabajos prácticos = T.P.
- *test* regional = T.R.

Se atribuye la misma importancia al *test* regional que a los otros dos elementos. De ahí la ponderación:

$$T.A. = 1/4$$

$$T.P. = 1/4$$

$$T.R. = 1/2$$

En la página siguiente tenemos el cuadro general de los grados y de los resultados después de la ponderación y el ajuste final.

¹⁴ Según D. Mather y colab., *op. cit.*, págs. 149-154.

ASIGNACION DEL GRADO FINAL

Alum- no N°	Test regional T.R.		Evaluación por la escuela		Ponderación			Total	Grado final
	Score	Note	T.A.	T.P.	T.R × 1/2	T.A × 1/4	T.P × 1/4		
1	59	2	3	3	1	1	3/4	2 ³ / ₄	3
2	77	1	1	1	1/2	1/4	1/4	1	1
3	51	3	4	3	1 1/2	1	3/4	3 1/4	3
4	12	6	5	5	3	1 1/4	1 1/4	5 1/2	6
5	53	3	3	4	1 1/2	1	1	3 1/2	4
6	40	4	2	5	2	3/4	1 1/4	4	4
7	66	1	2	2	1/2	1/2	1/2	1 1/2	1
8	60	2	3	4	1	3/4	1	2 3/4	3
9	38	5	4	5	2 1/2	1 1/4	1 1/4	5	5
10	70	1	2	2	1/2	1/2	1/2	1 1/2	1
11	56	2	2	3	1	1/2	3/4	2 1/4	2
12	69	1	2	2	1/2	1/2	1/2	1 1/2	1
13	44	4	4	5	2	1	1 1/4	4 1/4	4
14	64	2	1	2	1	1/4	1/2	1 3/4	2
15	19	6	5	5	3	1 1/4	1 1/4	5 1/2	6
16	49	3	2	4	1 1/2	3/4	1	3 1/4	3
17	54	3	2	1	1 1/2	1/2	1/4	2 1/4	2
18	47	4	1	2	2	1/4	1/2	2 3/4	3
19	52	3	3	1	1 1/2	3/4	1/4	2 1/2	2
20	48	3	3	3	1 1/2	3/4	3/4	3	3
21	50	3	1	2	1 1/2	1/4	1/2	2 1/4	2
22	24	6	3	4	3	3/4	1	4 3/4	5
23	61	2	4	2	1	1	1/2	2 1/2	2
24	57	2	1	3	1	1/4	3/4	2	2
25	42	4	3	5	2	3/4	1 1/4	4	4
26	35	5	5	2	2 1/2	1 1/4	3/4	4 1/2	4
27	45	4	4	1	2	1	1/4	3 1/4	3
28	41	4	2	4	2	3/4	1	3 3/4	4
29	27	5	5	5	2 1/2	1 1/4	1 1/4	5	5
30	43	4	3	2	2	1	1/2	3 1/2	4
31	67	1	1	1	1/2	1/4	1/4	1	1
32	31	5	5	3	2 1/2	1 1/4	1	4 3/4	5
33	72	1	2	2	1/2	1/2	1/2	1 1/2	1

Las notas ponderadas en bastardilla serán ajustadas en función del *test* regional (T.R.).

La lectura del cuadro nos muestra que el *alumno 1* obtiene las tres notas ponderadas: 1, 1 y $3/4$. O sea, en total, $2\ 3/4$, grado final¹⁵. El *alumno 3*: $1\ 1/2 + 3/4 = 3\ 1/4$; grado final 3.

¿Por qué se hace un último ajuste en el momento de asignar el grado final? Porque la suma de los grados produce una nueva concentración hacia la media; la contracción es tal que, en una escala final de 5 grados; toda la población corre el riesgo de encontrarse en el medio.

El siguiente ejemplo ficticio muestra ese fenómeno:

ALUMNOS	GRADO		
	Test R.	Trab. anual	Media
Pedro	1	5	3
Pablo	2	4	3
Juan	3	3	3
Rogelio	4	2	3
Enrique	5	1	3
Dispersión	4	4	0

Si volvemos al cuadro de la pág. 182 observamos que, entre los grados ponderados se encuentra un gran número de 1 (17 en total), mientras que en la columna *Total* se hallan solo dos. Las notas correspondientes a "sobresalientes" se perdieron en la suma de los grados.

El siguiente procedimiento asegura una mayor justicia distributiva entre los alumnos de una misma clase, y una mejor alineación en el nivel regional revelado por el *test* regional; éste es reconocido como la pauta más segura.

¹⁵ Las cifras en bastardilla en el cuadro indican que el moderador, de acuerdo con la escuela, ajustó la nota del T.A. con el propósito de armonizar con el nivel medio de la región.

Partimos del siguiente cuadro:

TEST REGIONAL		TOTAL DE GRADOS PONDERADOS		GRADO FINAL	
Grado (1)	Número de alumnos que obtuvieron ese grado (2)	Grado (3)	Número de alumnos que obtuvieron ese grado (4)	Número de alumnos a quienes se les asigna (5)	Grado (9)
1		1			1
		1 ¹ / ₄			
		1 ¹ / ₂			
		1 ³ / ₄			
2		2			2
		2 ¹ / ₄			
		2 ¹ / ₂			
		2 ³ / ₄			
3		3			3
		3 ¹ / ₄			
		3 ¹ / ₂			
		3 ³ / ₄			
4		4			4
		4 ¹ / ₄			
		4 ¹ / ₂			
		4 ³ / ₄			
5		5			5
Superior		Superior			No clasificado

Operaciones

1. Apuntar en la columna (2) el número de alumnos y calcular el total.
2. Empezar a apuntar en la columna (4) y detenerse cuando el total llega a ser igual al de la columna (2). En ese lugar se hace el primer corte. Para no poner en desventaja a

ciertos alumnos, se rebasa, si fuese necesario, el número de la columna (3), a fin de agotar el nivel donde nos hemos detenido.

Por ejemplo, veremos en el siguiente cuadro que se otorga el grado 2 a 7 alumnos y no a 6, porque 2 de ellos obtuvieron $2 \frac{1}{2}$. Es evidente que no se podía asignar un grado 2 a uno y 3 al otro.

TEST REGIONAL		TOTAL DE GRADOS PONDERADOS		GRADO FINAL	
Grado (1)	Número de alumnos que obtuvieron ese grado (2)	Grado (3)	Número de alumnos que obtuvieron ese grado (4)	Número de alumnos a quienes se les asigna (5)	Grado (6)
1	III 6	1	// 2	6	1
		1 1/4	/// 4		
		1 1/2			
		1 3/4			
2	III 6	2	/ 1	7	2
		2 1/4	/ 1		
		2 1/2	/// 3		
		2 3/4	// 2		
			/// 3		
3	III 7	3	/ 1	7	3
		3 1/4	/// 3		
		3 1/2	// 2		
		3 3/4	/ 1		
			// 2		
4	III 7	4	/ 1	7	4
		4 1/4	/ 1		
		4 1/2	// 2		
		4 3/4	/ 1		
			// 2		
5	IIII 4	5	// 2	4	5
Superior	III 3	Superior		2	No clasificado

Conclusión

Ninguno de los sistemas descritos es perfecto, pero todos permitirían mejorar nuestro sistema tradicional de calificación.

Una primera elección se realizará en función del objetivo perseguido: selección o comparación. Este segundo aspecto ha retenido fundamentalmente nuestra atención.

No abogamos a favor de un sistema en particular. La decisión pertenece a las autoridades políticas y pedagógicas y a los docentes.

Debe llevarse a cabo un intenso esfuerzo de experimentación en todos los niveles: pequeños grupos de profesores, investigaciones regionales y nacionales en colaboración con los centros psicomédico-sociales y los laboratorios de pedagogía experimental de las universidades.

En la medida en que esos trabajos avancen, es factible que veamos surgir un nuevo sistema que tenga en cuenta nuestras tradiciones y, además, los objetivos particulares de nuestro país. Sin embargo, casi de inmediato se hará sentir de nuevo la necesidad de poderosos centros regionales de investigación pedagógica. ¡Volveremos aún sobre este punto!

QUINTA PARTE

EL MITO DE LA CURVA DE GAUSS

EL PELIGROSO MITO DE LA CURVA DE GAUSS

En las ciencias humanas, la curva de Gauss, en forma de campana, desempeña un papel considerable porque es la imagen misma de la distribución de muchas aptitudes y cualidades: abundan los individuos término medio, mientras que los genios y los idiotas, los gigantes y los enanos resultan escasos.

La curva de Gauss es el reflejo de la ley del azar que preside nuestro nacimiento, o bien la resultante de la influencia de un gran número de factores que obran de manera más o menos independiente sobre un individuo o un objeto.

Como los *tests* miden a menudo aptitudes, rasgos de personalidad o realizaciones de vastas poblaciones, es natural que se distribuyan según el porcentaje gaussiano: de una manera general, 70% de regulares, 13% de buenos, 13% de mediocres, 2% de excelentes y 2% de malísimos.

Durante la elaboración de tales *tests*, se eliminan sobre todo las preguntas que serían contestadas por demasiados sujetos o por muy pocos. El fin perseguido estriba en clasificar a cada uno, en asignarle el lugar que le corresponde en un grupo numeroso. En suma, se trata de organizar una especie de concurso en el que el más fuerte ocupará necesariamente el primer lugar.

Por esta razón, muchos *tests* de aptitudes o inventarios de conocimientos constituyen excelentes *instrumentos de selección*.

En su clase, el educador persigue un objetivo completamente diferente. ¿No es su ideal que *todos* los alumnos aprendan a leer, a calcular y, de una manera general, a dominar a la perfección todos los conocimientos que se consideran necesarios o útiles para la sociedad? *Instruir no es seleccionar*. ¡Por el contrario! Significa hacer un esfuerzo para que *todos* tengan éxito. *Por lo tanto, significa luchar contra la curva de Gauss* tomada como modelo de selección.

Las consecuencias pedagógicas de estas observaciones son particularmente importantes.

CAPITULO 1

EVOLUCION DE LA CURVA DE CONOCIMIENTOS

En nuestro sistema de clases rígidas, cuando un maestro recibe, al empezar otro año lectivo, a un grupo de alumnos que no conoce, suele hallarse ante. . . dos curvas: una representa la distribución de aptitudes y otra la del saber.

1. La curva de aptitudes

En su acepción habitual —que conservamos por el momento— la palabra *aptitud* denota características, innatas o adquiridas, que se consideran como sintomáticas de la capacidad de un individuo para alcanzar un nivel de competencia más o menos elevado, en un determinado terreno.

En la enseñanza no estrechamente especializada (que para gran cantidad de nuestros alumnos llega más allá de los veinte años), la misma amplitud del caudal de conocimientos y de las capacidades por adquirir (matemáticas, idiomas, ciencias naturales, artes, etc.) hace imposible la selección muy rigurosa en función de una aptitud particular.

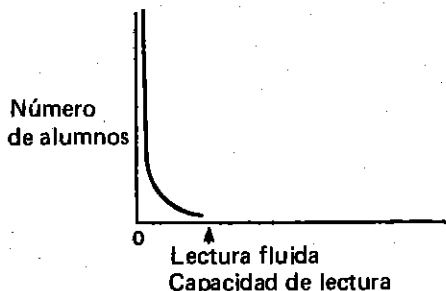
Por esta razón, hasta un nivel muy avanzado de la escolaridad, las aptitudes de los alumnos quedan distribuidas al azar.

En tales condiciones, un profesor de matemáticas, por ejemplo, que midiera las aptitudes especiales para su materia obtendría, normalmente, una distribución gaussiana o —dado el número probablemente reducido de alumnos— su esbozo.

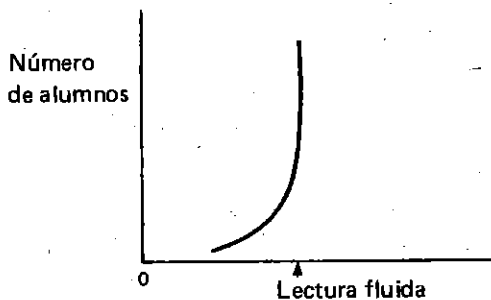
Por lo demás, los profesores apenas si tienen necesidad de recurrir a los *tests* para conocer esa situación; saben por

promover el aprendizaje de conocimientos *nuevos*. Imaginarse que todos los individuos que forman una clase posean exactamente la misma cantidad y calidad de conocimientos constituye, sin duda, una utopía. Sin embargo, el sistema de clases fijas que practicamos se basa en la hipótesis de que *todos* se encuentran aproximadamente en un mismo nivel. En caso contrario, ¿cómo nos atreveríamos aún a impartir la misma enseñanza a cada uno?

¿Y en la realidad? Tomemos el caso del ingreso en la escuela primaria. La mayoría de los niños no saben leer. Algunos marchan por buen camino; dos o tres ya leen de corrido. En ese momento, la curva de la capacidad de lectura sigue, aproximadamente, la forma de la letra i.

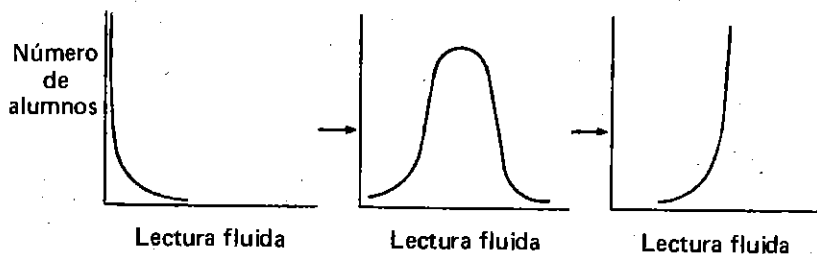


Ahora bien, aunque la curva de las aptitudes específicas para la lectura será muy probablemente gaussiana, el maestro no admite con facilidad que tan solo una parte de la división podrá aprender a leer. Mientras no se caiga en el perfeccionismo, puede afirmarse que después de un año o dos la gran mayoría de los alumnos leerán con fluidez. La curva de conocimientos habrá cambiado totalmente de forma: se parecerá a una j.



Entre la curva en i y la curva en j , es probable que haya existido un momento en el cual los más dotados adelantaran muy rápido, los menos dotados quedarán a la zaga y los regulares se situarán entre los dos grupos.

Desde el punto de vista esquemático, la evolución es, pues, la siguiente:



Capacidad en lectura

En el caso de la lectura, no se concibe que los padres o los responsables de la enseñanza puedan conformarse con una evolución distinta.

Pero, ¿a partir de qué nivel de la escolaridad tal exigencia pierde su validez?

CAPITULO 2

UNA PEDAGOGIA DE LA CURVA EN J

Un examen en el que los resultados se distribuyen según la curva en campana de Gauss, se presta adecuadamente a la selección, al concurso.

Ahora bien, debido a tres razones históricas principales, el espíritu de concurso ha impregnado nuestra enseñanza durante siglos:

1. Por razones socioeconómicas, solo una parte de la población por escolarizar tenía acceso a la escuela. Hace cien años, las familias modestas hacían la primera selección dejando terminar la escuela primaria tan solo a los más dotados de sus hijos, es decir, a aquellos que más éxito tenían en los exámenes. Las becas de estudios, parsimoniosamente distribuidas, se ganaban por concurso.
2. Hasta hace poco tiempo —antes de la era de la computadora— todos los países industrializados disponían de un enorme superávit de materia gris. Se explotaba, pues, aquella que se ofrecía a mejor precio y con más facilidad, es decir, existía la preocupación fundamental de identificar a los mejor dotados. Incluso para los niños afortunados, la enseñanza secundaria desempeñaba un papel selectivo.
3. Los conocimientos psicológicos y pedagógicos eran rudimentarios. Por esa razón, los maestros no estaban en condiciones de aplicar tratamientos sutiles a los alumnos con dificultades de aprendizaje.

Por lo demás, aún en la actualidad, a menudo se hace repetir un año al alumno que no ha sido capaz de dar los rendimientos mínimos exigidos para pasar de grado. Con

otras palabras, en lugar de aplicar remedios especiales nos conformamos con volver a colocar al alumno en las mismas condiciones (igual profesor, idéntico método) que lo llevaron al fracaso.

Puesto que cada año escolar se consideraba como un filtro para el siguiente, cierta cantidad de fracasos parecían, pues, normales. En suma, la distribución gaussiana se presentaba satisfactoria.

Pero la constante preocupación por promover a los más aptos determinó, en forma progresiva, una deformación pedagógica más sutil. Se llegó a creer que los conocimientos "que realmente cuentan", los conocimientos "profundizados", el "verdadero juego de las ideas abstractas", solo son asimilables para una minoría que posee aptitudes específicas en sumo grado. Y de acuerdo con este criterio se forjaron los métodos de enseñanza.

Por lo demás, los alumnos y sus padres aceptan sin mucha discusión ese estado de cosas. Pocas veces se inscribe gente en la sección "latín-matemáticas" que no se sienta particularmente dotada en ese sentido.

Es indudable que no se puede pretender hacer de cada uno, sin distinción, un virtuoso de las matemáticas, del piano o de la literatura. Pero, ¿dónde está el límite? ¿A partir de qué grado es imposible el acceso a un saber, de quienes poseen aptitudes intermedias o, incluso, mediocres?

Por naturaleza, cada uno de nosotros posee aptitudes nada más que mediocres en muchos terrenos. Por ejemplo, no faltan intelectuales más o menos refractarios a las matemáticas. Sin embargo, se observa con frecuencia que a fuerza de empeñarse, de perseverar, de hacerse repetir las explicaciones, de cambiar de manual o de maestro con el fin de hallar una forma de enseñanza conveniente, nociones matemáticas, antes consideradas como inaccesibles, llegan a ser dominadas perfectamente por algunos.

Con tal que le dediquen el tiempo necesario, los alumnos regulares (es decir, en vista de la selección ya hecha por los estudios anteriores, probablemente más del 80% de la población escolar de un determinado nivel) pueden llegar mucho más lejos de lo que suele imaginarse.

Desde este punto de vista, la nueva definición de aptitud,

propuesta por J. Carroll, es absolutamente válida: "La aptitud es la cantidad de tiempo que el estudiante requiere para dominar una materia."¹

El significado de tal concepción es considerable: si se les otorga el tiempo necesario, todos los alumnos que se encuentran normalmente en una clase podrían obtener un buen resultado, e incluso, muy bueno.

¹ J. Carroll, "A Model of School Learning", *Teachers College Record*, 1963, 64: pág. 723-733.

LA TEORIA DE LA EVALUACION FORMATIVA

La expresión *evaluación formativa* —que él opone a *evaluación sumativa*— fue acuñada por Michael Scriven¹.

La *evaluación sumativa* (S. Roller habla en ese caso de “norma psicotécnica”) nos es familiar. Para interpretar el puntaje obtenido en un *test* clásico de inventario de conocimientos o de inteligencia, se lo ubica en una distribución estadística: el rendimiento de un individuo se juzga en relación con el de otro. Asimismo, aún suele calificarse a los alumnos por comparación, de acuerdo con el orden creciente o decreciente de sus resultados escolares, y muchos padres aprecian el trabajo de sus hijos en función del lugar que llegan a ocupar.

Ahora bien, en ambos casos, el resultado es esencialmente relativo. Basta el hecho de que el grupo de referencia cambie su composición para que el resultado aparezca desde otra perspectiva.

Una simple diferencia de edad puede, asimismo, modificar considerablemente el aspecto de las cosas. Según muchas de las normas de los *tests* de conocimientos de uso corriente, un año más o uno menos de edad o de escolaridad bastan para que un mismo rendimiento se considere mediocre o bueno.

¹ M. Scriven, *The Methodology of Evaluation*, en R. Staker, ed., *Perspectives of Curriculum Evaluation*, Chicago, Rand McNally, 1967.

Véase asimismo: B. S. Bloom, “Learning for Mastery”, en *Evaluation Comment*, 1968, 2. Nuestra exposición del método de la evaluación formativa se basa directamente en esa publicación. Algunos resultados de experiencias en las que se aplicó la teoría del *mastery learning* fueron publicados por J. Block, B. Bloom, Hastings, 1970.

De modo, pues, que el hecho de haberse adquirido o no el conocimiento no era la preocupación primordial de los constructores de esos *tests*, sino más bien la velocidad con que se realizó esa adquisición. Remitiéndonos a la definición de J. Carroll, se evalúa, pues, la aptitud y no el contenido del aprendizaje.

Los partidarios de la *evaluación formativa* adoptan la actitud contraria a esa concepción.

Puesto que el alumno concurre a la escuela para aprender, ¿no es lo importante ubicarlo dentro de su ascensión al saber?

Imaginemos que un riguroso análisis científico nos revela que, con todos sus matices y complicaciones, la concordancia del participio pasado, empleado con el auxiliar haber, presenta cien casos diferentes. Para cada uno existe la posibilidad de definir criterios de dominio.

Puede fijarse entonces el número de casos por dominar, de acuerdo con el nivel escolar. En ese contexto, la evaluación escolar cambia de carácter.

Se califica al alumno en función de un criterio objetivo: el camino recorrido en la adquisición del saber.

Además, si aspiramos a conducir a todos los alumnos hasta un nivel mínimo de conocimientos de la noción o de la técnica, si no hasta su dominio total, la misma forma de enseñar se pone en tela de juicio: ya no es posible aplicar indistintamente a todos un solo método durante un mismo lapso.

¿En qué se convertiría la enseñanza dentro de esa perspectiva?

Se plantean tres problemas cruciales:

1. ¿Cómo jalonar la ascensión al saber?
2. ¿Cómo guiar al alumno?
3. ¿Dónde se ubica el límite práctico de esa pedagogía de la curva en *j*?

Según nuestro conocimiento, todavía no existen respuestas completas para estas preguntas. Semejante imprecisión del saber parece normal en ciencias naturales o en medicina. Tenemos que aprender a aceptarla también en las ciencias de la educación. Por esa razón, nuestra discusión desemboca más de una vez en la necesidad de emprender o continuar las investigaciones.

I. Jalonar la ascensión al saber

En un viaje, para determinar a qué distancia nos encontramos de la meta, dos condiciones tienen que cumplirse: por una parte, saber dónde estamos y adónde vamos; por la otra, disponer de un mapa que indique el camino con claridad.

En educación, asimismo, tenemos que definir los objetivos por alcanzar y determinar con precisión los aprendizajes particulares que nos conducirán hacia ellos. El problema varía según se trate, o no, de aprendizajes básicos. Para estos últimos hay que descubrir el encadenamiento "crítico" de las materias, es decir, aquel en que una de ellas solo es accesible si la anterior ha sido asimilada. Por esta razón, el aprendizaje de los conocimientos y de las técnicas de base ha de ser, en última instancia, lineal, mientras que las adquisiciones y las aplicaciones que van más allá ven abrirse frente a ellas caminos cada vez más numerosos.

Por ejemplo, sea cual fuere el método de enseñanza, no es posible aplicar completamente una regla de tres sin poseer —sobre todo— la noción de multiplicación y de división. En un nivel más elevado, ¿cómo se hará un cálculo integral sin saber qué es una función? Pero semejantes proposiciones resultan todavía demasiado vagas. ¿Cuáles son, exactamente, los aprendizajes imprescindibles y suficientes para hallarse en condiciones de asimilar la regla de tres? Y entre ellos, ¿cuáles son esenciales en relación con el resto?

La pregunta es temible. Y sería ingenuo creer que la reflexión y los tanteos que dieron por resultado la elaboración de nuestros mejores manuales escolares la hayan resuelto. Ciertamente, el buen encadenamiento de los capítulos constituye un comienzo de solución, pero basta con programar la menor noción, aunque más no sea según el sistema de Skinner, para descubrir toda la ignorancia y la imprecisión de los métodos de enseñanza actuales.

No obstante, es precisamente en los esfuerzos de programación emprendidos por un número cada vez mayor de investigadores y de docentes donde yace una de las grandes fuentes de progreso.

Sin duda, las ciencias deductivas constituyen el terreno de elección del método que esbozamos. Aunque debiéramos limitarnos a ellas, no sería despreciable la obra realizada. Pero el

límite no aparece tan aceleradamente. Ahí donde la inducción predomina en la elaboración del saber, la deducción reaparece para su aplicación. Por lo demás, el aprendizaje inductivo puede asimismo guiarse en forma sistemática.

II. Guiar al alumno

Una vez definido con claridad el camino que conduce a un aprendizaje, hay que encauzar al alumno. En cuanto a las adquisiciones fundamentales decidimos, además, que *todos* deberían, en principio, llegar a la meta.

En ese caso es esencial la frecuente evaluación de los adelantos. De ahí proviene la necesidad de *ejercicios de dominio* y de *tests diagnósticos* referidos a materias muy limitadas y administrados por los maestros mismos.

Esos instrumentos —que solo la colaboración entre docentes e investigadores permitirá construir en número suficiente— indican *dónde se encuentra el alumno* (lo que le otorga la nota) y *dónde experimenta dificultades*.

Los *remedios* son múltiples y han de entrar en nuestra práctica pedagógica de todos los días:

- a) Indicar con precisión la parte del curso por repasar.
- b) El trabajo en subgrupos: si un grupo de más de tres alumnos encuentra una o varias dificultades afines, el profesor tiene la ventaja de dedicar parte del tiempo de la clase a ese grupo, mientras los demás trabajan independientemente.
- c) Si el grupo cuenta con un máximo de tres alumnos, B. Bloom sugiere que cada uno de ellos sea integrado en un grupo heterogéneo de dos o de tres. Como los alumnos ya no se hallan en competencia, sino que luchan por dominar una materia, la ayuda mutua los favorece.
- d) En escuelas que cuentan con varios profesores de una misma materia, un alumno debería tener la libertad de pedir una explicación, e incluso una breve lección particular, a otro profesor.
- e) Variar la presentación: poner a disposición de los alumnos manuales que expliquen la noción de diversas maneras, permitir el acceso a la enseñanza programada, a los medios audiovisuales; variar el nivel de abstracción.
- f) En general, no fijar para todos el mismo tiempo para las mismas cosas.

Semejante práctica rinde beneficios. B. Bloom informa sobre una experiencia realizada según esa línea, donde más del 80% de los alumnos obtuvieron la mejor nota al finalizar el año. Representa el triunfo de la pedagogía de la curva en *j*.

¿Podrá ser siempre así? Nada permite afirmarlo.

III. Límite práctico del sistema

Incluso en el caso de que todos los alumnos de nuestras escuelas sean capaces de alcanzar un alto grado de dominio siempre que se les otorgue el tiempo suficiente, resulta evidente que más allá de cierto límite la duración se vuelve prohibitiva.

¿Cómo permitir que trascurren veinte años, en vez de seis, en la enseñanza secundaria! Sin embargo, esta comprobación no ha de poner en tela de juicio el sistema. Lo principal consiste en orientar a los alumnos en función de sus aptitudes reales. El ciclo de observación y orientación, introducido de manera progresiva en la enseñanza secundaria de la mayoría de los países, ofrece esa posibilidad.

Después de esa selección, cuando el estudiante es admitido en una clase, debe ser capaz de aprender en forma normal lo esencial de *todo* lo que allí se enseña.

IV. El sistema de enseñanza puesto en tela de juicio

La pedagogía de la curva en *j* llama a la nivelación hacia arriba; sin embargo, sería una quimera imaginar que permite la nivelación hacia el . . . genio.

Tal como hemos visto, menos del cinco por ciento de la población posee aptitudes *excepcionales*. Estas deben desarrollarse al máximo, tanto por respeto a la persona como en el interés de la comunidad.

Un nuevo sistema de enseñanza permite, a la vez, instruir a cada uno de acuerdo con un ritmo conveniente, y da la oportunidad de desenvolverse a los superdotados: se trata de la escuela sin clases, donde se trabaja en grupos homogeneizados según las aptitudes específicas *pour une seule branche*²

² El sistema no ha de confundirse con el *streaming*, que consiste en la elaboración de clases homogéneas con ayuda de *tests* de aptitudes generales o, peor todavía, en función de los resultados escolares globales. Se

En las condiciones actuales, la clase constituye una picota. ¿Por qué un alumno de seis años, que sabe ya leer cuando ingresa en primer grado, no podría participar en los ejercicios de lectura del año siguiente? ¿Por qué un alumno de cuarto año secundario no podría seguir un curso de matemáticas del nivel correspondiente a un tercer o segundo año?

En nuestras viejas escuelas primarias de campaña, donde un solo maestro enseñaba a veces a los seis grados, tales desplazamientos eran comunes. En los colegios secundarios que, actualmente, cuentan con varias divisiones de la misma edad, es posible trabajar en cuatro o cinco niveles diferentes de aptitudes, por lo menos en algunas materias principales. Un sistema similar funciona perfectamente, desde hace muchos años, en escuelas tales como la New Trier High School, que describimos detalladamente ya en 1961³.

Refutemos por anticipado la objeción financiera. Una escuela con grupos homogéneos específicos no cuesta más que otra; por el contrario. Si se agregan el ahorro de tiempo y talento y la disminución de los fracasos que el sistema permite, puede afirmarse, sin riesgo de equivocarse, que el beneficio es considerable. . .

comete, de esta manera, un doble error científico: en la gran mayoría de los casos, un sujeto está dotado en forma desigual para las distintas materias del programa; además, homogeneizar de una manera global hace descender el rendimiento del conjunto. Por otra parte, en ese sistema se llega a menudo a una semisegregación social. Todos esos inconvenientes fueron tan claramente señalados que en Suecia el *streaming* está prohibido por la ley.

³ "L'individualisation de l'enseignement dans une école multilatérale", en *Education*, may. 1961.

picota.
cuando
recios
cuarto
cas del

donde
tales
darios
misma
erentes
es. Un
muchos
l, que

Una
ás que
npo y
istema
que el

CONCLUSIONES Y RECOMENDACIONES

oría de
stintas
al hace
ma se
nientes
hibido

ltilaté-

Expusimos conclusiones parciales en la medida en que avanzamos en este estudio ya prolongado. Sin embargo, estamos lejos de haber agotado el tema y, más aún, de haber aportado soluciones definitivas.

¡Salir de la rutina irreflexiva y pretender la traducción de las grandes teorías pedagógicas a la práctica cotidiana de la educación, nos hace descubrir tantos escollos, tantas condiciones por cumplir, que casi nos invade la duda y el desánimo!

Sin embargo, aun corriendo el riesgo de dejar insatisfechos a nuestros lectores, no hemos querido ocultarles la dificultad y menos todavía hacerles creer en la existencia de una docimología consumada, capaz de llevar a una evaluación perfecta si se le proporcionan los medios. Probablemente, y por suerte, la medición rigurosa de los rasgos humanos más sutiles resultará siempre imposible: ¡que el destino nos proteja de las máquinas para pesar el alma!

El hecho de que la evaluación continua se incorpore de una manera funcional en el proceso de enseñanza y de aprendizaje, del que no puede desligarse, no elimina la necesidad de una docimasia, es decir, de una técnica del examen y la calificación, o para expresarlo en otros términos, de una técnica de observación sistemática del rendimiento inmediato o diferido.

Por lo demás, la exigencia de evaluación comparada, que el educador puede deplorar por razones ideales, sobrepasa el ámbito escolar: corresponde a un rasgo fundamental de nuestra civilización.

Tal vez esto se modifique algún día. Si nuestra cultura sigue intelectualizándose y, por consiguiente, tiende a aumentar su complejidad, no vemos, sin embargo, cómo podría renunciar del todo a los exámenes y concursos.

Nuestra posición docimológica es resueltamente ecléctica en su deseo de conciliar las ventajas indudables de nuevas técnicas, de índole sobre todo estadística, con el aporte innegable y fundamental del análisis cualitativo.

Para limitarnos a un solo ejemplo de tal eclecticismo, no podemos concebir que las decisiones de promoción que intervienen al final de cada año de estudios universitarios se tomen únicamente en función de los resultados obtenidos en un *test* de elección múltiple que se presta para la calificación automática. La introducción de ese tipo de pruebas es altamente deseable, sin duda, para el control objetivo y profundizado de los

conocimientos, pero no basta. Debe agregarse la evaluación del trabajo del año, y el encuentro último entre el maestro y su alumno ha de aportar todos los matices que el enfoque cuantitativo ha dejado en la penumbra.

Así como la introducción de las máquinas en la industria permitió a la inteligencia imponerse a la fuerza muscular y a la rutina envilecedora, así también el control automático del conocimiento de los hechos, métodos y técnicas posibilita un examen final centrado realmente sobre los procesos mentales y los rasgos de personalidad más nobles. ¿Qué importancia relativa atribuiremos a los tres tipos de evaluación? En este caso, la decisión corresponde a la comunidad académica, esclarecida por las investigaciones docimológicas.

Parece imponerse cierta cantidad de *recomendaciones*.

Precisamente por nuestra ignorancia y la gran complejidad de la evaluación es urgente, en primer lugar, que una *red de centros de investigación en educación* cubra todo el país y se ponga al servicio del sistema escolar. Las funciones de esos centros se pusieron de relieve en el trascurso de nuestro análisis: investigaciones vinculadas con los problemas fundamentales y desarrollo de los métodos, las técnicas y los instrumentos necesarios. Esos centros no deben injertarse como cuerpos extraños en el conjunto de nuestras instituciones educativas, sino trabajar en estrecha colaboración.

Y más específicamente, *cada universidad* debería crear en su seno un *departamento de estudios de los problemas relacionados con la enseñanza y los exámenes*. Resulta inconcebible que la institución científica por excelencia excluya del rigor de su análisis a una de sus dos razones de ser: la educación.

La *definición de los objetivos* generales y especiales, fundamento de todo el edificio de la enseñanza y de la evaluación, requiere el trabajo mancomunado de un gran número de investigadores y de docentes en ejercicio de todas las disciplinas. *Comisiones especializadas* posibilitarán la cooperación imprescindible y la comunicación —en ambos sentidos— entre la base y la cúspide de la jerarquía pedagógica.

La reforma de los exámenes y de los modos de evaluación hará desaparecer, sin duda definitivamente, el cálculo del porcentaje, difundido casi en forma universal en nuestras escuelas hasta el momento actual. Sean cuales fueren sus deficiencias, la escala de porcentajes ofrecía, sin embargo, la

ventaja de ser familiar a todos, y cada uno tenía por lo menos la ilusión de comprenderla. Actualmente se experimenta en nuestras escuelas con las escalas más diversas. *Una unificación del sistema de calificación y una campaña de información* resultan necesarias sin más pérdida de tiempo. De lo contrario, la comunicación entre la escuela y la familia, entre las distintas escuelas y entre éstas y las autoridades será difícil, si no imposible. Sin embargo, antes de tomar una decisión cuya importancia es indudable, debería escucharse a los expertos en docimología y, de una manera más general, a los expertos en mediciones en ciencias humanas.

Se realizará, asimismo, un amplio esfuerzo de información dirigido a los docentes llamados a cambiar profundamente sus hábitos de evaluación. El medio más eficaz estriba, en apariencia, en la *organización de experimentos locales* enderezados hacia la preparación de los exámenes, la calificación y la moderación. Esos experimentos se incorporarían, por ejemplo, en los trabajos necesarios para la introducción de la enseñanza secundaria renovada, en Bélgica, y en la modificación de la formación de maestros; de manera gradual, todas las escuelas participarían.

Con el tiempo se constituirán *comisiones de moderación locales y regionales*.

Una acción demasiado precipitada podría poner en peligro una renovación sumamente loable. En particular, se evitará sobre todo decepcionar o engañar a los docentes tratando de ponerlos bajo la autoridad de docimólogos improvisados.

Obra humana por excelencia, la educación solo mejorará su calidad en la medida en que los educadores acepten en profundidad las ideas nuevas y traten de superarse a sí mismos.

APENDICES

**ESTUDIO COMPARADO DE UNA PREGUNTA DE EXAMEN
PRESENTADA SEGUN EL METODO TRADICIONAL
Y EL DE ELECCION MULTIPLE**

*Presentación de la pregunta*¹ (método tradicional).

1. a) Dar la lista de las causas de hipoglucemia.
- b) Describir las lesiones anatomopatológicas resultantes de la hipoglucemia.

Para facilitarse cierto grado de uniformidad en la apreciación, los encargados de la corrección confeccionaron en conjunto la siguiente planilla:

- A. Igual valor para ambas partes de la pregunta.
- B. Para obtener una nota de 75/100, el candidato debe haber indicado en la lista las causas de:

- 1) Exceso de insulina.
- 2) Tumor funcional de los islotes de Langerhans.
- 3) Influencia endocrina por hipofuncionamiento de la hipófisis y las suprarrenales.
- 4) Enfermedad del hígado que impide la reserva del glucógeno (necrosis aguda) o su liberación (enfermedad de Von Gierke).

2. ¿En qué se reconoce la diferencia entre las lesiones causadas por un solo ataque agudo de hipoglucemia y las causadas por una hipoglucemia crónica?

3. Describir las alteraciones visibles a nivel cerebral.

- C. Para obtener una nota entre 75 y 90/100, el candidato debe haber indicado:

- Una lista que muestre que ha comprendido el mecanismo de acción de cada una de las causas.
- O bien una lista que abarque otras causas, pero indicando que comprende que los mecanismos de

¹ Moore, Robert A., *Method of Examining Students in Medicine*, en "Journal of Medical Education", en. 1954, vol. 29, n° 1.

Traducción tomada del informe francés sobre la reforma de los exámenes de medicina, *op. cit.*

homeostasis tienden a corregir la hipoglucemia en casos tales como:

- 1) Ingesta disminuida.
- 2) Absorción perturbada.
- 3) Metabolismo aumentado, como en el caso de hipertiroidismo.
- 4) Infección grave.
- 5) Agotamiento físico.
- 6) Glucosuria renal.

D. Para obtener una nota superior a 90/100, el candidato debe haber indicado:

- Una lista lógica de las cuatro causas mayores y de las seis causas menores, además de un tipo idiopático.
- Una diferenciación de la hipoglucemia aguda y crónica mediante una descripción detallada de las lesiones cerebrales.

Presentación de la pregunta (método de preguntas de elección múltiple).

He aquí cómo podría presentarse la misma pregunta según el método nuevo.

1. Se puede determinar si el candidato es capaz de reconocer las cuatro causas más importantes y diferenciarlas de las menos importantes en la siguiente pregunta:

INSTRUCCIONES. Cuatro de las cinco frases numeradas son comunes a uno de los tres trastornos funcionales de la lista alfabética (A, B, C).

Indique cuál es la excepción y el trastorno funcional común a las cuatro frases restantes.

- | | |
|---|---|
| 1. Exceso de insulina | A. Hipoglucemia clínicamente comprobable |
| 2. Tumor funcional de los islotes de Langerhans | B. Hiperglucemia clínicamente comprobable |
| 3. Glucosuria renal | C. Glucosuria clínicamente comprobable |
| 4. Hipofunción hipofisaria | |
| 5. Enfermedad de Von Gierke | |

Si el candidato responde correctamente, muestra saber que

1, 2, 4 y 5 pueden producir una hipoglucemia clínicamente comprobable, que esto no es así en el caso 3 y que ninguna combinación de los cuatro casos entre los cinco puede asociarse a la hiperglucemia ni a la glucosuria. Con otras palabras, está probada la posesión de conocimientos positivos y negativos.

2. Si queremos conocer cuánto sabe el candidato acerca de los fenómenos que rigen el equilibrio del nivel de glucosa sanguínea y de insulina, podemos formular la pregunta siguiente:

INSTRUCCIONES. Cada una de las siguientes frases se compone de dos partes: la exposición de un hecho y su razón de ser.

Indique para cada una de las frases numeradas la letra A, B, C, D o E, según que:

- A. La exposición del hecho y su razón sean verdaderas y mantengan una relación causa-efecto.
 - B. La exposición del hecho y su razón sean verdaderas, pero no guarden relación de causa a efecto.
 - C. La exposición del hecho sea verdadera, pero la razón resulte falsa.
 - D. La exposición del hecho sea falsa, pero la razón constituya un hecho o principio aceptado.
 - E. Tanto la exposición del hecho como la razón resulten falsas.
1. El nivel de glucosa en la sangre cae brutalmente después de una hepatectomía, porque el glucógeno contenido en el hígado constituye la fuente principal de la glucosa.
(A)
 2. El diagnóstico anatomopatológico de un adenoma de los islotes de Langerhans en el páncreas implica que la enfermedad era hiperinsulínica, porque todos los adenomas de los islotes son funcionales y segregan insulina.
(E)
 3. Los enfermos que sufren hipertiroidismo tienen todas las probabilidades de gestar una hipoglucemia, porque existe un hiperinsulinismo asociado.
(C)
 4. Los pacientes que sufren la enfermedad de Von Gierke presentan una disminución del volumen de
(C)

glucosa sanguínea porque, en esa enfermedad, el hígado no hace reserva de glucógeno.

5. Una hipoglucemia que data de varios meses no trae secuelas, porque las alteraciones celulares producidas por la hipoglucemia son reversibles.
- (E)

Si el candidato responde correctamente a esta serie, eso muestra que no desconoce:

- a) Que el glucógeno hepático constituye la fuente principal que permite el mantenimiento del volumen de glucosa en la sangre.
 - b) Que todos los tumores de los islotes de Langerhans no son funcionales.
 - c) Que los enfermos de hipertiroidismo sufren hipoglucemia, pero que la causa no es un hiperinsulinismo asociado.
 - d) Que aquellos que padecen la enfermedad de Von Gierke tienen hipoglucemia, pero que ésta no se debe al hecho de que no haya glucógeno en el hígado.
 - e) Que la hipoglucemia crónica trae secuelas y que las alteraciones celulares resultantes no son reversibles.
3. Existe la posibilidad de determinar si el candidato posee algunas nociones simples respecto de las alteraciones a nivel cerebral, mediante la siguiente pregunta:

INSTRUCCIONES. Cada una de las exposiciones incompletas (numeradas) va seguida de cinco complementos para elegir. Indique, en cada caso, el complemento más conveniente.

1. Las alteraciones anatómicas secundarias de una hipoglucemia crónica son puestas en evidencia con mayor frecuencia en:
 - a) el bazo
 - b) el riñón
 - c) las suprarrenales
 - d) el cerebro
 - e) la tiroides

2. Entre las alteraciones provocadas por la hipoglucemia crónica a nivel cerebral, las más importantes afectan:

- a) las neuronas
 - b) los astrocitos
 - c) las células endoteliales
 - d) las células oligodendríticas
 - e) las células microgliales
3. Entre las alteraciones de las células nerviosas provocadas por la hipoglucemia aguda, la más evidente es:
- a) el desplazamiento del núcleo
 - b) la desaparición de la pared celular
 - c) la alteración de los cuerpos de Nissl
 - d) la rotura del núcleo celular
 - e) la fusión de las mitocondrias
4. Entre las alteraciones provocadas por la hipoglucemia crónica a nivel cerebral, la más significativa es:
- a) la hidrocefalia interna
 - b) la fibrosis de la aracnoides
 - c) la destrucción de las neuronas
 - d) la proliferación de los astrocitos
 - e) la proliferación de las células endoteliales

Si el candidato responde correctamente a esta serie, muestra que sabe que las principales alteraciones de la hipoglucemia crónica alcanzan las neuronas del cerebro, que la hipoglucemia aguda provoca una alteración de los cuerpos de Nissl en las neuronas y que la hipoglucemia crónica ocasiona la destrucción de células nerviosas.

4. Es factible determinar si el candidato comprende el mecanismo de los trastornos, mediante la siguiente pregunta:

INSTRUCCIONES. En la lista alfabética están indicados cinco mecanismos diferentes que pueden conducir a la hipoglucemia.

Anotar la letra adecuada después de cada una de las proposiciones numeradas, junto con el mecanismo responsable del trastorno.

- A. Aumento de la actividad metabólica.
- B. Hiperinsulinismo.
- C. Acumulación anormal de glucógeno en el hígado.
- D. Falta de reserva de glucógeno en el hígado.
- E. Hipofunción de la hipófisis o de las suprarrenales.

- (B) 1. Adenomas de los islotes de Langerhans.
- (A) 2. Ejercicio físico violento.
- (A) 3. Hipertiroidismo.
- (E) 4. Enfermedad de Simmonds.
- (C) 5. Enfermedad de Von Gierke.
- (D) 6. Hepatitis epidémica.
- (C) 7. Enfermedad de Addison.

Si el candidato responde correctamente a esta serie, esto muestra que comprende las bases del metabolismo de los glúcidos y conoce los factores que lo influyen.

Recapitulación

De esta manera, por medio de diecisiete "preguntas objetivas", pusimos en evidencia todos los conocimientos requeridos para otorgar una nota superior a 90/100, a saber:

1. Lista de las cuatro causas principales.
2. Conocimiento de la diferencia entre los efectos de la hipoglucemia crónica y la aguda.
3. Descripción de las alteraciones a nivel cerebral.
4. Comprensión de los mecanismos.
5. Lista de las causas menores.
6. Diferenciación de las lesiones causadas por la hipoglucemia aguda y la crónica.

REDACCION EN LENGUA MATERNA
INGLATERRA – GENERAL CERTIFICATE
OF EDUCATION 1967
FIN DEL CICLO SECUNDARIO – NIVEL COMUN

PRUEBA I. Dos preguntas – 1½ hora.

1. Elija uno de los temas siguientes. Dedíquele aproximadamente 1 hora.
 - a. Una noche de niebla.
 - b. Un comerciante compra y revende, el mismo día, un objeto de ocasión. Describa ambas escenas.
 - c. ¿Cuál sería su política si usted dirigiera los programas de radio o televisión?
 - d. Se dispersa una muchedumbre. Describa la escena.
 - e. El placer de la fotografía, o del dibujo, o de la danza, o del ciclismo.
 - f. ¿Cree usted que varones y niñas tienen las mismas oportunidades de hacer carrera?
 - g. ¿Cuáles son sus reacciones frente a los progresos y realizaciones de la exploración espacial?

2. Elija uno de los siguientes temas. Dedíquele aproximadamente 1/2 hora.
 - a. Describa un entrenamiento destinado a mejorar su rendimiento en un deporte de su elección.
 - b. Después de la visita efectuada a una empresa, redacte un informe para sus compañeros acerca de las condiciones de trabajo y las perspectivas de porvenir que ofrece.
 - c. Describa con claridad *uno* de los siguientes aparatos y explique cómo funciona: un "walkie-talkie", un secador de cabello, una aspiradora, una licuadora.
 - d. En la ciudad que usted habita existe la idea de crear un centro comercial donde estará prohibida la circulación de vehículos. Escriba una carta a un diario local exponiendo sus opiniones acerca del proyecto.

PRUEBA II. Cuatro preguntas — 1¼ hora.

1. Haga un resumen del siguiente pasaje, en buena prosa continua y con un total de 110 palabras como máximo. Al final, indique cuántas palabras empleó. El pasaje tiene 314 palabras. *Viene a continuación un texto sobre el exceso de confianza en la ciencia y en la tecnología.*
2. Lea el siguiente pasaje; luego, conteste las preguntas.
(El texto describe dos grandes tipos de excursionistas —los que siguen a un guía y los que se entregan a la aventura— y subraya el interés de las excursiones geológicas.)
 - a. Explique con sus propias palabras la diferencia entre los dos tipos de excursionistas.
 - b. ¿Qué consejo da el autor respecto de las grutas? Fórmúlelo con sus propias palabras.
 - c. Explique brevemente por qué el autor cree que la geología es: un *hobby* divertido — un *hobby* instructivo.
 - d. El autor explica que emplea la expresión “cavar una zanja” en sentido literal. Aclare por qué el sentido es, en este caso, literal.
 - e. Explique las siguientes expresiones: precauciones prescritas; ser perfectamente consciente de la naturaleza de sus actos.
 - f. Elija cuatro de las siguientes palabras. Sustitúyalas por sinónimos o perífrasis que podrían utilizarse en el texto sin cambiar el sentido (. . .).
3. Responda, a elección, a una de las dos preguntas siguientes:
 - a. Elija tres palabras entre las que presentamos. Construya oraciones (6 en total) que muestren que esos términos pueden emplearse en dos sentidos diferentes (. . .).
 - b. Defina en una frase tres de las siguientes palabras: monopolio - interludio - prólogo - microscopio - antídoto.
4. Responda, a elección, a una de las dos preguntas siguientes:
 - a. Explique clara, pero brevemente, la diferencia de senti-

do entre cada par de palabras (se refiere a *could-should*; *can-may*; *will-shall*; *might-must*).

- b. Vuelva a escribir correctamente el siguiente pasaje, respetando todas las ideas. Puede cambiar la expresión, el orden de las palabras e ideas, la ortografía y la puntuación.

(Sigue un texto plagado de errores y que abarca un centenar de palabras.)

“Por ejemplo si se proyectara hacer pasar una ruta a través de una ciudad pero que una casa histórica estuviese en el camino entonces los planes deberían cambiarse, causando gastos considerables, para rodear el edificio creando un desvío en la ruta y haciéndola así peligrosa para los autos. El desvío de esa ruta no solo causa erogaciones sino que aumenta el precio de varias cosas especialmente si el desvío es grande porque si la distancia agregada es, digamos, de diez kilómetros y un camión llevando ciertos artículos recorre la ruta, el camión consumiría entonces más combustible necesitaría más tiempo para llegar y su vida se acortaría en diez kilómetros.”

CONSIGNAS PARA LA CORRECCION

PRUEBA I — Máximo 50 puntos.

El esquema de calificación que figura más adelante no constituye sino una guía preliminar. Podrán hacerse agregados y enmiendas en la reunión de examinadores que se convocará después de una primera lectura de los trabajos.

Pregunta 1 (máximo, 35 puntos).

Se espera un mínimo de 400 palabras, pero las redacciones no han de calificarse principalmente en función de su extensión. Hay que tener en cuenta el tema elegido y la manera de tratarlo. Aun si es breve, una redacción cuya argumentación es concisa y la expresión adecuada debe obtener más puntos que una larga narración informe.

El examinador debería tener una idea clara acerca de lo que es una redacción que recibe la nota mínima necesaria para pasar (16 puntos). Tal redacción ha de contener ideas razonables, pero no muy originales. La expresión ha de ser clara, pero sin ninguna distinción en particular. En el trabajo no deben encontrarse más que unos pocos errores maquinales. Los candidatos que hayan rebasado ese nivel general serán recompensados, y los que permanecieron por debajo sufrirán una sanción.

En el documento adjunto se encuentran notas detalladas acerca de las cualidades por observar en las redacciones. Sin embargo, se cuenta con que los examinadores calificarán en función de su impresión general y que, por lo tanto, no asignarán un número fijo de puntos para los distintos aspectos. Si el tema se presta a la controversia, las ideas y su encadenamiento pueden ser más importantes que en una redacción descriptiva, donde el vocabulario ocuparía un lugar más amplio. Confiamos en el juicio de los examinadores.

Al reservar 35 puntos a esa primera pregunta, se ha querido destacar su mayor importancia en el conjunto de los exámenes. Muchos candidatos son mediocres. Sin embargo, si el examinador no dispersa sus notas ampliamente, esa primera pregunta carecerá del peso suficiente en el total de los resultados.

Grupo A (29-35 puntos).

La redacción es de calidad excepcional.

Grupo B (22-28 puntos).

La calidad del trabajo es superior al promedio.

Grupo C (14-21 puntos).

El trabajo es de calidad regular.

Grupo D (7-13 puntos)

El trabajo es de un nivel inferior al que se considera satisfactorio.

Grupo E (0-6 puntos)

El candidato es incapaz de presentar sus ideas en forma coherente.

Esa prueba es un examen en lengua materna. En la prueba 1, examinamos la capacidad del candidato para expresar sus opiniones, experiencias, impresiones, sentimientos e intereses. No se trata de un *test* de conocimientos generales ni de una evaluación de las aptitudes del candidato. Cuando una joven describe con claridad una escena en inglés, ha de obtener una nota favorable si los materiales utilizados son pertinentes, aun si el examinador cree que la joven aborda el tema de una manera demasiado sentimental. Por lo demás, si el examinador tiene la impresión de que detrás de la masa de prosa incoherente se hallan emociones profundas y elevadas actitudes morales, no se ha de limitar a ese último aspecto: cuenta solo lo que está escrito. Para rendir satisfactoriamente la prueba, son esenciales la claridad de expresión y la precisión de estilo.

Pregunta 1 (máximo, 35 puntos).

- (a) Se aceptan tanto una narración como una descripción. En el primer caso, la noche de niebla debe desempeñar un papel esencial.
- (b) En términos generales, las dos escenas deben estar equilibradas. Recompensar la vivacidad de la narración, del diálogo, de la descripción y el contraste.
- (c) Se espera una definición clara de la política acerca de la cual deben desarrollarse varios puntos. El candidato tiene que haber elegido la radio o la televisión; no puede reunir las dos.
- (d) El tema del trabajo es la dispersión; sin embargo, se autoriza una breve introducción.

- (e) Una respuesta coherente, clara, de moderada extensión es mejor que largas divagaciones y repeticiones.
- (f) El tema no es fácil; hay que recompensar generosamente el adecuado ordenamiento de los argumentos y los ejemplos bien elegidos.
- (g) Un tratamiento puramente narrativo no puede aceptarse, pero algunos ejemplos de progresos realizados tal vez resulten imprescindibles para explicar las reacciones.

Pregunta 2 (máximo, 15 puntos).

- (a) Se espera una descripción clara y lógica.
- (b) El informe ha de tratar los tres aspectos de la cuestión. Se insiste, sobre todo, en la información ofrecida; por lo tanto, no se exigirá una forma particular de informe.
- (c) Se exige una descripción del objeto y, asimismo, una explicación de su funcionamiento. No castigar severamente los errores materiales; calificar simplemente la claridad de expresión.
- (d) Pueden encàrarse numerosos aspectos. De esta manera, se sancionará toda idea inadecuada.

Descontar:

1 punto por redactar mal la dirección.

1 punto por falta de coherencia entre el encabezamiento y los saludos finales.

1 punto si el alumno ha firmado "Sr. John Smith" o "Señorita Jeannette Smith".

1/2 punto por otros errores de disposición o de puntuación, o de ortografía en las palabras esenciales.

Todos esos temas proporcionan suficiente material para media hora de trabajo. Se espera un mínimo de 200 palabras. La elección es abundante.

PRUEBA II — Máximo, 50 puntos.

Pregunta 1 (máximo, 16 puntos).

Los puntos atribuidos a esta pregunta suelen ser mucho menores que para las otras. Se pide a los examinadores que no consideren 11 como tope máximo para el resumen.

A.- Atribuir un máximo de 2 puntos a cada uno de los siguientes aspectos. Para obtener 2 puntos, el candidato debe

haber comprendido con toda claridad la idea y expresarla con corrección. Matizar las notas con $1\frac{1}{2}$, 1 ó $\frac{1}{2}$ punto. Al final del trabajo, cuente una vez más el total en función de la fluidez y de la coherencia del resumen completo. Si un pasaje es incoherente, suprime a lo sumo $\frac{1}{2}$ de los puntos otorgados; si resulta más bien duro, reste $1/8$. Tal corrección debería aparecer en la redacción anotándose, por ejemplo, $10 - 1 = 9$. Toda sustracción de puntos destinada a sancionar un texto que excede el límite fijado o una proposición no solo inexacta, sino absurda, también debe indicarse por separado. El total final se marcará con un círculo al margen.

1. En la actualidad, el hombre de la calle acepta los descubrimientos científicos,
2. sin dudar de su origen, validez o efectos.
- 3 y 4. La exigencia de nuevos progresos destinados a elevar el nivel de vida no se detiene nunca.
5. La gente tiene confianza en el hombre de ciencia y en su trabajo,
6. y cree que no se puede frenar el progreso.
7. Aunque reconoce que los hombres de ciencia no siempre están de acuerdo acerca de las seguridades brindadas por los nuevos descubrimientos,
8. el público está convencido de que terminarán por hallar la unanimidad o, por lo menos, un amplio acuerdo.

B.- El límite de 120 palabras deja un margen generoso. Suprime un punto por cada grupo de 5 palabras que sobrepase el máximo. No incluya en el recuento palabras de introducción tales como "En este pasaje, el autor explica que...". Vuelva a contar las palabras. No acepte sin ratificarlo el número anotado por el candidato.

Pregunta 2 (máximo, 20 puntos).

Al calificar esta pregunta, otorgue el máximo en cada sección al candidato que expone con claridad el tema.

- (a) (1) Aquellos que quieren que todo se prevea por ellos (1)
- (2) y los que prefieren tener simplemente un plan general que les permita seguir la inspiración del momento (espíritu de exploración) (2)

- (b) Visitar las grutas. (1)
 A menos que usted esté guiado por alguien que
 conoce muy bien el terreno. (1)
 Respete todas las reglas de seguridad. (2)
- (c) (1) Casi no se requiere un equipo especial. (2)
 (2) El geólogo hace continuamente nuevos descubri-
 mientos. (1½)
 en pequeña escala. (½)
- (d) Cuando el geólogo parte una piedra, realiza, de
 hecho, un acto único, pues nadie podrá volver a
 partir esa piedra. (2)
- (e) (1) Medidas de seguridad (1)
 impuestas o recomendadas. (1)
 (2) Para comprender exactamente (1)
 las consecuencias de lo que uno hace. (1)
- (f) Subterráneo:
 lo que está debajo del suelo, de la superficie de
 la tierra; (1)
 bajo tierra. (1)
- Espeleólogos:
 personas que estudian científicamente las caver-
 nas; exploradores de cavernas. (1)
- Concienzudamente:
 de buena fe, sin caer en distracciones, con
 mucha aplicación (1)
 honestamente (½)
 fielmente (½)
- Etcétera.

Para las subpreguntas (a), (b) y (e) no conceda ningún punto a los alumnos que se limiten a copiar parte del texto.

Pregunta 3 (máximo, 6 puntos).

- (a) Asigne un punto para toda oración construida correcta-
 mente.
 No otorgue punto alguno en caso de construcción inco-
 rrecta.
- (b) Conceda un punto por cada definición exacta expresada
 en una oración correcta (máximo, 3 puntos).
 Asigne un punto por cada oración en la que el sentido de
 la palabra que comienza con el mismo prefijo aparezca
 con claridad (máximo, 3 puntos).

$\frac{1}{2}$ punto por una definición expresada en una oración incorrecta.

$\frac{1}{2}$ punto por una definición dada en una oración incompleta.

Ningún punto, si la palabra que comienza con el mismo prefijo no se presenta en una oración.

Un *monopolio* es una propiedad exclusiva en manos de una firma (o)

Un *monopolio* es el nombre otorgado a una firma que posee derechos comerciales exclusivos.

Un *interludio* es un intervalo introducido en el desarrollo de una obra (o un intervalo en el desarrollo de un acontecimiento).

Etcétera.

Pregunta 4 (máximo, 8 puntos).

- (a) Asigne un punto por oración.

Verifique la presencia de las siguientes ideas. . .

No otorgue punto alguno si no se ve con claridad a qué oración se refiere el candidato.

- (b) Los candidatos responderán de diferentes maneras.

Reduzca 1 punto por falta de ortografía, 2 puntos por toda construcción defectuosa, 1 punto por una expresión incorrecta, $\frac{1}{2}$ punto por la omisión o el empleo erróneo de una coma *esencial*, 1 punto por cada idea omitida.

Revea la nota total en función de la impresión general del pasaje. .

Para las preguntas 1 y 2, si el candidato obtiene la mitad de los puntos, redondear hacia la unidad superior (ejemplo: $6\frac{1}{2} = 7$).

Si las dos preguntas dan una nota con $\frac{1}{2}$ punto, redondear una hacia arriba y la otra hacia abajo.

Iguales observaciones pueden hacerse con respecto a las preguntas 3 y 4.

EJEMPLO DE ENSEÑANZA SEMIINDIVIDUALIZADA¹

La New Trier Township High School, Winnetka

La New Trier Township High School es un gran colegio secundario superior, con enseñanza semiindividualizada, donde el alumno puede corregir su orientación hasta el fin de la adolescencia.

Admite alumnos de 14 años cumplidos² y goza de una gran reputación, tanto en Estados Unidos como en el extranjero.

Siguen sus cursos en forma regular 3740 estudiantes³. El cuerpo docente se compone de 222 integrantes, distribuidos de la siguiente manera:

40	profesores de inglés
31	” de matemáticas
30	” de educación física
29	” de idiomas extranjeros
27	” de “Social Studies” (historia, geografía, instrucción cívica, sociología y economía)
17	” de ciencias
9	” de música
7	” de cursos técnicos
7	” de comercio
6	” de arte dramático
6	” de pintura y dibujo
6	” de conducción de automóviles
4	” de economía doméstica

¹ Extracto de un artículo que publicamos en *Education*, may. 1961.

² Es sabido que la organización de la enseñanza en Estados Unidos no es uniforme. La N. T. High School se basa en el sistema “NK 8-4”: un año de Nursery School, un año de jardín de infantes, 8 años de primaria y 4 años de secundaria. Los otros dos sistemas más frecuentes son: NK 6-3-3 y NK 6-6. Existen, asimismo, NK 7-5, NK 6-2-4 y NK 6-4-4.

³ Esos números se refieren al año lectivo 1959-1960. Fueron tomados de: *Information for College Admission Officers*, N. T. Township High School, nov. 1959, o bien se recogieron en el lugar.

- 2 profesores de higiene
- 1 profesor de enseñanza especial (retardados mentales educables).

Se cuenta con un profesor por cada 15-16 estudiantes, y las divisiones suelen tener aproximadamente 25 alumnos.

Los servicios de un profesor que no asume alguna responsabilidad especial (como jefe de un departamento, por ejemplo) comprenden 24 períodos de 40 minutos por semana: 4 cursos de 5 períodos y 4 períodos de *counseling* (asesoramiento).

El nivel intelectual de los estudiantes, juzgado sobre la base de los *tests* clásicos, es elevado. Aproximadamente un 80% de los que terminan sus estudios en la New Trier obtienen resultados superiores a la media nacional en los *tests* de aptitudes y de conocimientos (SCAT, STEP, *National Merit*). Un 92% de los diplomados siguen estudios superiores.

La individualización de los programas

El principio fundamental de la labor pedagógica de la New Trier High School está definido ya en las primeras líneas de su programa de estudios: "Nadie puede creer en la dignidad de los hombres sin experimentar un profundo placer frente al espectáculo de su infinita variedad."⁴

Impregnados de esta manera del espíritu del Plan de Winnetka, los estudios están organizados de modo de brindar a cada uno la posibilidad de desarrollarse según su ritmo propio, según sus capacidades e inclinaciones. Esto no significa de manera alguna que se admitan la fantasía estéril o las soluciones fáciles.

De acuerdo con el sistema difundido en todo el país, el estudiante sólo obtiene su diploma de fin de estudios si, en el transcurso de éstos, ha logrado un total de puntos o *créditos* fijado por el consejo de administración de la escuela, que se basa, a su vez, en normas generales. Esos *créditos* expresan, en una unidad convencional, la importancia cualitativa y cuantitativa de los distintos cursos por un semestre (5 *créditos* corresponden, por ejemplo, a un semestre de curso "menor",

⁴ New Trier Township High School, *Curriculum Guide*, dic. 1959, pág. 1.

a razón de cinco períodos por semana). El estudiante sólo adquiere *créditos* si obtiene una nota superior a un límite mínimo estipulado.

Para obtener el certificado de estudios de la New Trier High School hay que haber ganado 350 *créditos* en 4 años.

El juego de los *créditos* constituye a la vez una fuente de flexibilidad y de seguridad, porque permite delimitar con exactitud el campo de libertad del alumno.

En principio, éste establece su programa de estudios tal como se compone la minuta de una comida, eligiendo de la carta aquello que más atrae y conviene.

Aquí tenemos, por ejemplo, la lista de los cursos ofrecidos para primer año, con indicación de los *créditos* que otorgan si se siguen con éxito durante un semestre⁵:

Cursos "mayores"

(Otorgan 10 *créditos* por semestre; salvo otra indicación, tienen que seguirse durante 2 semestres consecutivos.)

Inglés	Idiomas extranjeros
Algebra	Matemáticas generales
Alimentación (1 semestre)	Pintura-dibujo (2 períodos por día)
Instrucción cívica	Fotografía
Comercio	Radioaficionado
Dibujo Industrial	Trabajar la madera
Electricidad (1 semestre)	Trabajos manuales
Historia universal	Corte y confección (1 semestre)
Idiomas extranjeros	

Cursos "menores"

(Número de *créditos* entre paréntesis)

Arte dramático	(6)	Dicción	(6)
Canto coral	(3)	Armonía	(5)
Dactilografía	(5)	Orquesta sinfónica	(5)
Curso de armonía	(6)	Trabajos manuales (1 período por día)	(5)
Dibujo industrial (1 período por día)	(5)		

⁵ Cf. New Trier Township High School, *Registration Bulletin for Freshmen, 1959-1960. Planning a Course of Study*, pág. 2.

Sin embargo, varias restricciones influyen en la elección del alumno. En primer lugar, ciertos cursos, considerados indispensables para la cultura básica de todos los integrantes de la nación, son obligatorios:

Curso	Duración obligatoria	Créditos
Lengua materna	4 años	80
Matemáticas	2 años	40
"Social Studies"	2 años ⁶	40
Ciencias	1 año	20
Educación física	4 años	16
Automóvil: teoría y conducción	1 semestre	3

Total 199

Aproximadamente 2/3 del total de *créditos* exigidos van a parar, en consecuencia, a los cursos obligatorios. Sin embargo, según veremos, esas exigencias pueden satisfacerse de una manera muy libre; en efecto, hemos comprobado la existencia de unas diez posibilidades diferentes que permiten cumplir válidamente el año de ciencias que se requiere.

Otro factor importante guía al estudiante en la elaboración de su programa: la profesión o los estudios superiores a que aspira. En particular, cada universidad determina sus condiciones de admisión y especifica, sobre todo, cuántos *créditos* tiene que haber adquirido el alumno en la enseñanza secundaria para determinadas materias.

Por último, los padres, maestros, consejeros pedagógicos y orientadores hacen lo posible por guiar al estudiante de acuerdo con lo que mejor conviene a sus intereses y velan para que obtenga el mayor provecho de sus potencialidades.

Como la semana escolar tiene 5 días de 8 períodos de 40 minutos —sin contar la reunión cotidiana con el consejero pedagógico (20') y el almuerzo, que siempre se toma en el colegio (25')—, suele proponerse al alumno que fije un programa medianamente cargado que le deje tiempo para el estudio personal. Aquí tenemos cuatro ejemplos típicos para primer año⁷:

⁶ Un año dedicado en forma obligatoria a la historia de Estados Unidos.

⁷ New Trier Township High School, *Courses for Freshmen*, abr. 1959, pág. 4.

I		II		III		IV	
Curso	per. sem.	Curso	per. sem.	Curso	per. sem.	Curso	per. sem.
Inglés	5	Inglés	5	Inglés	5	Inglés	5
Algebra	5	Algebra	5	Algebra	5	Algebra	5
Latín	5	Ciencias	5	Historia	5	Historia	5
Pint./dib.	10	Comercio	5	Aliment.	7	Ciencias	5
Gimn.(niñas)	4	Gimn. (varones)	5	Canto	3	Idiom.extr.	5
	—	Música instrum.	5	Gimn.(niñas)	4	Gimn.(var.)	5
	29		—		29		30
Estudio personal	11		30		11		10
	40		40		40		40

Se observará cuánto contrastan esos planes de trabajo con el desperdicio de esfuerzo tan frecuente entre nosotros. Prácticamente, todos los cursos académicos pueden concentrarse en la mañana.

Alrededor de una cuarta parte del tiempo se reserva a los estudios personales. Las horas que así quedan libres se destinan a menudo a trabajos de investigación en la biblioteca del colegio, que en Estados Unidos desempeña un papel incomparablemente más importante que en Bélgica.

La individualización de la enseñanza

No solo elige el alumno las materias que más le convienen, sino que, además, la enseñanza de cada una de ellas se adapta a sus posibilidades.

En efecto, todas las materias importantes pueden estudiarse en cinco niveles distintos de aptitudes: inferior, promedio débil, normal, acelerado y avanzado. De esa manera, el esfuerzo exigido al estudiante guarda siempre proporción con sus posibilidades, posea o no una inteligencia superior, esté poco dotado, o bien resulte fuerte en un campo y atrasado en otro.

Como el alumno pocas veces sigue todos sus cursos en un mismo nivel, el peligro de una segregación general, según las aptitudes, parece mínimo; por lo demás, el colegio está siempre muy atento a ese problema y reagrupa sistemáticamente a todos los estudiantes en oportunidad de ciertas

actividades. Los esfuerzos se concentran, asimismo, en inculcar una verdadera tolerancia frente al talento más o menos grande de los compañeros de estudios y en alentar a cada uno para lograr su superación: "En el deporte, no todos son capaces de sobresalir; lo mismo sucede con los estudios. Pero se espera que cada uno brinde lo mejor de sí..."⁸.

El sistema de enseñanza en distintos niveles se practica en New Trier desde el fin de la primera Guerra Mundial, con el beneplácito general. Los estudiantes se distribuyen, término medio, según los siguientes porcentajes:

— superiores	15-20%
— normales (regulares fuertes)	40-55%
— regulares débiles	36-40%
— limitados	5- 8%
— avanzados (<i>seniors</i>)	±10%

Cuando los alumnos entran en la New Trier High School, los grupos se establecen provisionalmente sobre la base de los resultados escolares anteriores y de otros informes reunidos de acuerdo con un sistema que estudiaremos más adelante. A partir de entonces, los resultados obtenidos en el mismo colegio corregirán y guiarán las asignaciones que nunca son definitivas, sea cual fuere la materia.

Semejante flexibilidad requiere, ciertamente, una organización compleja; mas no debemos exagerar esta dificultad.

En el cuadro siguiente aparecen todas las posibilidades ofrecidas en el segundo semestre del año lectivo 1959-1960. Algunas notas al margen aclaran el espíritu de los estudios⁹.

Cada curso está simbolizado por un número de tres guarismos:

- El guarismo de las centenas indica el año en que se da: tendremos, pues, 1, 2, 3 y 4.
- El guarismo de las decenas indica el semestre.

⁸ New Trier High School, *Guide Book to New Trier*, Winnetka, 1959, pág. 39.

⁹ Se utilizaron los dos documentos de base siguientes: New Trier Township High School, *Curriculum Guide*, 1959, y *Program of Classes*, 1959-1960.

- 1 = primer semestre
 2 = segundo semestre
 0 = puede seguirse en el 1^{er} o 2^o semestre
 3 = curso de vacaciones de verano.

c) El guarismo de las unidades señala el nivel de aptitud:

- 1 = nivel inferior
 2 = regular débil
 3 = normal
 4 = enseñanza acelerada
 5 = avanzado
 6 = seminarios
 9 = todos los niveles reunidos.

Materias y niveles de cursos	Observaciones
<p><i>Lengua materna</i></p> <p>125</p> <p>124</p> <p>123</p> <p>122</p> <p>122 R</p> <p>121</p>	<ul style="list-style-type: none"> - Para estudiantes de inteligencia superior, sobresalientes en lengua materna (comprende también biología e historia). - Curso enriquecido; basado en las obras literarias de nivel universal. - Curso medio. - Curso medio inferior; orientado en particular hacia los alumnos que aún no adquirieron un método racional de trabajo. - R = "remedial". Destinado a los alumnos identificados por la oficina de <i>testing</i> como poseedores de una marcada deficiencia en lectura (comprensión y rapidez) y en ortografía. - Reúne a los alumnos que presentan una deficiencia general en todas las ramas de lengua materna.
<p>224 - 223 - 222 - 221</p> <p>324 - 323 - 322 - 321</p> <p>424 - 424+ - 423 - 422 - 421</p>	<p>Las diferencias de nivel en 2^o, 3^{er} y 4^o año corren <i>mutatis mutandis</i> paralelas a las indicadas anteriormente respecto a 1^{er} año.</p> <p>323 = periodismo.</p> <p>424(+) = "Great Books".</p>
<p><i>Matemáticas</i></p> <p>Algebra A (acelerado) 124</p> <p>Algebra - 124</p> <p>Algebra E (experimental) 124</p> <p>Algebra - 123</p> <p>Algebra E 123</p>	<p>La diferencia entre los niveles 2, 3 y 4 no reside tanto en la aceleración como en el método de enseñanza y la penetración.</p> <p><i>En el nivel 2</i>, las explicaciones son detalladas; muchas aplicaciones; no se</p>

Materias y niveles de cursos	Observaciones
Algebra — 122	hace incursión alguna en campos vecinos.
Algebra E 122	
Matemáticas de base 121	
Matemáticas profundizadas (20 créditos) 225	<i>En el nivel 3</i> , el estudiante debe trabajar más por sí mismo: teoría más rigurosa, enriquecimiento.
Geometría 224	
Geometría 223	<i>En el nivel 4</i> , los conceptos se tratan rápidamente; estudio más profundo; numerosas incursiones en campos vecinos.
Geometría 222	
Matemáticas 222	Para estos 3 niveles se utiliza un mismo manual, y todos los alumnos estudian el mismo capítulo al mismo tiempo; esto permite pasar en cualquier momento de un nivel a otro.
Matemáticas de base 221	
Matemáticas profundizadas (20 créditos) 315	
Matemáticas 324	4 semestres de matemáticas son obligatorios; sin embargo, 2 de ellos pueden dedicarse a contabilidad o bien a aritmética comercial.
Algebra 323	
Algebra 322	
Geometría 303	
Matemáticas profundizadas 425	
Matemáticas 424	
Algebra (nivel universitario) 403	
Trigonometría 403	
Uso de la regla de cálculos 323	
(2 créditos; va junto con álgebra 322)	
<i>"Social studies"</i>	4 semestres de "Social studies" son obligatorios; 2 de ellos han de dedicarse a la historia de Estados Unidos.
Instrucción cívica 123, 122, 121	
<i>Historia universal</i> 125, 124, 123	
Historia: tiempos modernos 224	
Historia: antigüedad 204	
Historia: Edad Media 202, 203, 204	
Historia: universal 221	
Geografía: 222 - 223	
Historia: tiempos modernos 323	
Historia: Gran Bretaña 303	
Historia: EE.UU. 324-323-322-321	
Historia: EE.UU. (profundizada) 425	
Historia: contemporánea (Europa) 425	
Historia: Gran Bretaña (profund.) 404	
Historia: América latina 403	
Historia: Extremo Oriente 403	
Historia: universal, siglo XX 402	
Historia: EE.UU., siglo XX 402	
Instrucción cívica 403	
Sociología 403	
Ciencias económicas 403	

<i>Materias y niveles de cursos</i>	<i>Observaciones</i>
Ciencias Biología 123-124-125 Radioaficionado 123 Biología 224-223-222 Química 225 Biología 323-322-321 Electrónica 323 Física 325 Química 324-323 Química (seminario) 426 Física 425-424-423 Ciencias 409	Los alumnos que deseen continuar estudios en las facultades de ciencias de las universidades, son invitados a seguir 6 semestres de ciencias y 8 de matemáticas.
Idiomas Latín 124-123-122 224-223-222 324-323 424-423 Alemán 124-123 224-223-223C 324-323 426 Ruso 124 224 324 Francés 125-124-123 224C 224 223C 223 324C 324 323C 323 324= 424C 423C Español 125 124 123 122 224 224C 223 223C 222 324 324 = 323 322 424 423	Ningún curso de idioma es obligatorio. Sin embargo, se anima fuertemente a los alumnos a seguirlos si piensan dedicarse a estudios superiores. Se cree que, muy pronto, las universidades exigirán para el ingreso que el alumno haya estudiado por lo menos un idioma extranjero durante 3 años. C indica que se trata de la continuación de un curso anterior. = curso de conversación.
Artes Dibujo-pintura 129 Trabajo manual 121 Dibujo-pintura 229C 229 Trabajo manual 221 Dibujo-pintura 329 Cerámica 329 Joyería 329 Historia del arte 329 Cerámica 429 Joyería 429 Pintura-dibujo 429	Todos estos cursos pueden seguirse a razón de 2 períodos por día (= curso "mayor": 10 <i>créditos</i>) o de 1 período por día (= curso "menor": 5 <i>créditos</i>). Sin embargo, Historia del arte 329 es siempre "mayor".
Comercio Comercio general 122 123	Objetivos perseguidos por este curso: 1. Preparación para estudios superiores de ciencias económicas.

<i>Materias y niveles de cursos</i>		<i>Observaciones</i>
Comercio general	222	2. Formación general.
Contabilidad	223	3. Preparación para los estudiantes que deseen trabajar por horas en el comercio durante sus estudios universitarios.
El punto de vista del consumidor	323	4. Preparación profesional para los alumnos que no realizarán estudios superiores.
Venta	302	5. Posibilidad para los estudiantes que buscan su camino, a fin de determinar si la carrera comercial podría interesarles.
Publicidad	303	El 422 = una reunión de 30' en la escuela todas las mañanas; el resto consiste en trabajos prácticos (20 h/semana).
Organización comercial	303	
Taquigrafía	323	
Dactilografía	303	
Práctica de oficina	322	
Secretaría	423	
Taquigrafía	423	
Derecho comercial	423 403	
Práctica comercial	422	
<i>Automóvil</i>		Una ley promulgada en 1955 hace obligatorio este curso en todos los colegios secundarios de Illinois (3 créditos).
Teoría y práctica (conducción)	209	
<i>Economía doméstica</i>		Destinado fundamentalmente a las jóvenes; sin embargo:
Alimentación	109	- los muchachos interesados en hotelería pueden seguir los cursos de alimentación;
Vestido	109	- los muchachos interesados en arquitectura cuentan con la posibilidad de seguir los cursos de decoración de interiores.
Alimentación	209	
Vestido	209	
Decoración de interiores	329	
<i>Cursos técnicos</i>		- No se aspira a una formación profesional, sino a convertir al estudiante en "un consumidor inteligente de los productos industriales" (pág. 62).
Dibujo industrial	123 122	- Cada uno de estos cursos confiere 5 ó 10 créditos, según el tiempo dedicado a ellos.
Madera	123 121	- Las universidades reconocen esos créditos y desean, cada vez más, que los estudiantes adquieran una formación técnica.
Electricidad	103	
Dibujo (arquitectura)	223	
Madera	222	
Metalurgia general	223	
Dibujo (mecánica)	323	
Metalurgia general	323	
Nociones de arquitectura	323	
Lectura de planos	302	
<i>Música</i>		Se trata de una iniciación para el aficionado o bien de estudios profundos que pueden preparar para las carreras musicales.
Canto	129	
Orquesta	129	
Curso de armonía	129	
Apreciación musical	129	

<i>Materias y niveles de cursos</i>		<i>Observaciones</i>
Canto	229	
Pequeño grupo vocal	229	
Apreciación musical	229	
<hr/>		
Redacción	329	
<hr/>		
Madrigales (grupo vocal)	429	
Redacción	429	
<hr/>		
Coral (selección de 32 estud.)	029	
Opera	029	
Gran coral	029	
Armonía (novatos-grupo de último año-grupo de honor)	029	
Orquesta (<i>seniors</i>)	029	
Conjunto instrumental	029	
Piano	029	

Educación física - Deportes - Danza

Oratoria - Arte dramático

Fotografía

Cruz Roja (primeros auxilios)

El estudio de cada alumno antes de su ingreso en la New Trier

Hemos afirmado que la escuela recoge detallados informes acerca de sus futuros alumnos para que éstos encuentren de inmediato un lugar, por lo menos provisional, en el complejo sistema que acabamos de describir. Veamos cómo se procede.

La mayoría de los estudiantes que frecuentan la New Trier provienen de seis escuelas públicas y de siete escuelas confesionales de los alrededores.

A partir del mes de enero, los servicios de *testing* de la *High School** entregan a las seis escuelas públicas *tests* de inteligencia, de aptitudes y de conocimientos para administrarlos a los niños que dieron a conocer su intención de inscribirse en los cursos de New Trier. Los alumnos de las

* En inglés en el original.

escuelas confesionales llegan a Winnetka en abril para someterse ahí a pruebas similares.

Por otra parte, cada escuela debe llenar una ficha individual que contiene, además de los resultados de los *tests* mencionados, un cuadro que resume la apreciación general de los profesores¹⁰:

APRECIACIONES GLOBALES

Inteligencia	1	2	3	4	5
Aplicación	1	2	3	4	5
Sentido de la responsabilidad	1	2	3	4	5
Conducta en la escuela	1	2	3	4	5
Cualidades de jefe	1	2	3	4	5

APRECIACIONES ESPECIALES

Inglés	1	2	3	4	5
Matemáticas:					
Razonamiento	1	2	3	4	5
Conocimientos básicos	1	2	3	4	5
"Social Studies"	1	2	3	4	5
Ciencias	1	2	3	4	5

Código: 1 = sobresaliente; 2 = superior al promedio; 3 = regular; 4 = inferior al promedio; 5 = deficiente. Se marca el número correspondiente.

Después de la administración de los *tests*, el director del servicio de *counseling* (asesoramiento) de New Trier se reúne con los consejeros de las distintas escuelas primarias para recoger informaciones complementarias concernientes no solo a la instrucción, sino también a la historia, la familia, la salud de los futuros estudiantes.

Luego, éstos son entrevistados en su escuela. La conversación se refiere, en primer término, a sus aptitudes (artes, deportes, habilidades diversas) y sus gustos (materias preferidas), y después, a la familia; el entrevistador trata de descubrir eventuales problemas. La entrevista termina con una pregunta "choque".

En el mes de mayo, padres y alumnos —que ya tuvieron oportunidad de estudiar el programa de los cursos— son

¹⁰ Cf. New Trier High School, *Test and Personal Data Card*.

invitados a asistir a una reunión de información en la *High School*. Se reciben las inscripciones en los distintos cursos.

La dirección de New Trier posee entonces suficientes elementos como para constituir grupos de aproximadamente 30 estudiantes que, durante todos sus estudios, serán atendidos por el mismo asesor pedagógico con quien se reúnen todas las mañanas. En esos grupos se busca sistemáticamente heterogeneidad, tanto con respecto a las aptitudes como a los orígenes socioeconómicos. Los sexos, sin embargo, están separados, y el asesor es siempre del mismo sexo que sus estudiantes. Además, se procura conferir una personalidad equilibrada a cada grupo evitando, sobre todo, que se junten demasiados adolescentes con problemas.

El día en que se inician las clases, no solo están organizados los cursos y horarios, sino que cada estudiante sabe exactamente lo que tiene que hacer: no ignora, por ejemplo, que para tal materia tiene que reunirse con el grupo de los regulares y, para tal otra, con el grupo superior; sabe que un asesor lo espera, e incluso, qué libros, cuadernos y equipo de deportes debe comprar, y se le ha comunicado de antemano el precio exacto de cada uno de esos artículos.

Señalemos, por último, que aproximadamente 15 días después de iniciadas las clases se invita a los padres a una primera reunión, en la que se encontrarán con el asesor de sus hijos y se les explicarán las razones de la clasificación en los distintos niveles, el significado de las manifestaciones extraescolares, etcétera.

Una semana después, padres y madres vienen a la escuela junto con su hijo y comparten con él todo un día de clase (en esa ocasión, la jornada abarca desde las primeras horas de la tarde hasta la noche). No es difícil imaginar cuánto favorece tal experiencia la comprensión entre la escuela y la familia.

El estudiante, que ha comenzado sus estudios secundarios superiores en tan excelentes condiciones, será seguido día tras día por su asesor.

Los consejeros pedagógicos

Gracias a esos consejeros pedagógicos, la New Trier High School logra esa verdadera hazaña que consiste no solo en guiar con eficiencia a cada uno de sus 3.740 estudiantes por el

laberinto de los programas, sino también en ayudarlos tanto en el plano psicológico como en el social y médico.

El *counseling* (asesoramiento) se halla en manos de:

126 asesores pedagógicos con dedicación parcial profesores que dedican 1/5 de su tiempo a esa misión;

10 asesores con dedicación exclusiva que dirigen a los anteriores;

6 psicólogos con dedicación exclusiva.

Durante la cotidiana reunión matutina, cada grupo de 30 estudiantes cuya constitución ya señalamos se encuentra con su asesor para discutir problemas de disciplina, de carrera y, en general, de todo lo que puede interesar a la mayoría. La atmósfera relajada de esas reuniones constituye una excelente transición hacia el arduo trabajo que pronto comenzará.

El estudiante puede, asimismo, consultar individualmente con su asesor en cualquier momento.

Además, durante el año, el consejero hará por lo menos una visita personal a la familia de cada alumno que pertenece a su grupo.

Está sobrentendido que misión tan delicada no se improvisa; requiere hombres abiertos a los problemas de la juventud y especialmente formados para ayudarla.

Al reclutar a los profesores, el consejo de administración del colegio atribuye una importancia muy particular a las cualidades intelectuales y morales indispensables para el apadrinamiento de los estudiantes, y otorga preferencia a quienes las posean.

Durante el primer año de su función, el consejero tiene que asistir a 50 reuniones de formación de aproximadamente 1/2 hora cada una, y después de cada ciclo de 4 años se le somete a un nuevo entrenamiento.

La preparación básica se desarrolla de la siguiente manera:

I. *Antes de iniciarse las clases.*

Un cierto número de reuniones se destinan, en primer término, a estudiar la administración del colegio. Luego, el consejero recibe toda la información recogida sobre sus futuros alumnos.

Por último, algunos días antes de terminar las vacaciones el director de estudios de primer año reúne a los "*freshmen helpers*", es decir, a los estudiantes de último año que

ayudarán al consejero en su tarea. Durante las primeras 9 semanas, el asistente tomará parte en todas las reuniones matinales. En el transcurso de las 9 semanas siguientes, solo prestará servicio dos veces por semana; después, intervendrá únicamente a pedido del consejero.

II. *Durante el año lectivo.*

Con la presidencia de un director especializado, la formación sigue por medio de charlas familiares entre los directores de los departamentos (inglés, matemáticas, ciencias, etc.), los jefes de los distintos servicios (directores de *testing*, psicólogo, bibliotecario, médico-asistente social, delegado ante la asociación de padres) y los adultos responsables de los clubes estudiantiles.

El jefe de consejeros convoca, asimismo, a ciertas reuniones especiales para encarar problemas nuevos concernientes a la admisión de alumnos, la revisión de programas, etcétera.

Por último, en abril, una conferencia de clausura permite realizar el balance de las actividades del año que toca a su fin y encarar las primeras previsiones para el siguiente.

BIBLIOGRAFIA

- Agazzi, A., *Les aspects pédagogiques des examens*, Estrasburgo, Consejo de Europa, C.C.C., 1967.
- Bacher, F., "La normalisation de la notation", en *Docimologie et Education*, número especial de la revista "Les Sciences de l'Education", n° 2-3, 1969, 131-156.
- Bacher, F., "L'évaluation des résultats scolaires au niveau de l'école moyenne", en *Le Travail humain*, 1965, 28, 219-230.
- Bazin, R., "Les Français d'aujourd'hui et leurs examens", en *Education et Gestion*, 4, 1970, 3-10.
- Block, J., *Mastery Learning*, Nueva York, Holt, Rinehart y Winston, 1970.
- Bloom, B., Hasting y Madaus, *Formative and Summative Evaluation of Student Learning*, Nueva York, Mc Graw-Hill, 1970.
- Bonnardel, R., "Application de la méthode d'analyse factorielle de Thurstone à l'étude de la notation des copies d'examens", en *Le Travail humain*, VIII, 1946, 130-139.
- Britton, J., "Experimental Marking of English Composition Written by Fifteen-Year-Olds", en *Educational Review*, Birmingham, vol. 16.1, 1963, 17-23.
- Britton, J., Martin, N. y Rosen, H., *Multiple Marking of English Composition, An Account of and Experiment*, Londres, H.M.S.O., 1966.
- Bruce, G., *Secondary School Examinations. Facts and Commentary*, Oxford, Pergamon Press, 1969.
- Brunelle, L., *Pourquoi des examens?* París, Société des Editions rationalistes, 1968.
- Choppin, B. y Purves, A., "A comparison of open-ended and multiple choice items dealing with literary understanding", en *Research in the Teaching of English*, 3, 1, 1969, 15-24.
- Demangeon, M. y Larcebeau, S., "Une expérience de correction multiple", en *BINOP*, 1958, 14, 131-156.
- Docimologie et Education*, número especial de la revista *Les sciences de l'Education*, 2-3, 1969, 166 págs.
- Edgeworth, F. V., "The Statistics of Examinations", en *Journal of the Royal Stat. Society*, 1888, 51, 599-635.
- Examen des examens*, número especial de *Cahiers de pédagogie*, 92, set. 1970.
- Examinations Bulletins*, Londres, H.M.S.O.
- N° 1. *The Certificate of secondary education: some suggestions for teachers and examiners*, 1963.
- N° 2. *The C.S.E.: Experimental examinations - Mathematics*, 1964.
- N° 3. *The C.S.E.: An introduction to some techniques of examining*, 1964.

- N° 4. *The C.S.E.: An introduction to objective-types examinations*, 1964.
- N° 5. *The C.S.E.: School-Based examinations*, 1965.
- N° 6. *The C.S.E.: Experimental examinations: Technical drawing*, 1965.
- N° 7. *The C.S.E.: Experimental examinations - Mathematics 2*, 1965.
- N° 8. *The C.S.E.: Experimental examinations: Science*, 1965.
- N° 9. *The C.S.E.: Trial examinations: Home economics*, 1966.
- N° 10. *The C.S.E.: Experimental examinations: Music*, 1966.
- N° 11. *The C.S.E.: Trial examinations - Oral English*, 1966.
- N° 12. *Multiple marking of English compositions*, 1966.
- N° 13. *The C.S.E.: Trial examinations: Handicraft*, 1966.
- N° 14. *The C.S.E.: Trial Examinations - Geography*, 1966.
- N° 15. *Teachers' experience of school based examining (English and Physics)*, 1967.
- N° 16. *The C.S.E.: Trial examinations - Written English*, 1967.
- N° 17. *The C.S.E.: Trial examinations - Religious knowledge*, 1967.
- N° 18. *The C.S.E.: The place of the personal topic - History*, 1968.
- N° 19. *C.S.E.: Practical work in science*, 1969.
- N° 20. *C.S.E.: A Group Study Approach to Research and Development*, Londres, Evans-Methuen, 1970.
- Fischer, H., "Wechselwirkungen zwischen Unterrichtszielen, Didaktik und Prüfungen", en *Eidgenössischen Technischen Hochschulen Bulletin*, Zurich, ag. 1970, 9-14.
- French, J. W., *Schools of Thought in Judging Excellence in English Themes*, Princeton, E. T. S., 1961.
- Hartog, P. y Rhodes, E. C., *An Examination of Examinations*, Londres, McMillan, 1936.
- Hartog, P., *The Marking of English Essays*, Londres, McMillan, 1941.
- Hinton, E. M., *An Analytical Study of the Qualities of Style and Rhetoric Found in English Compositions*, Nueva York, 1940.
- Hotyat, F., *Les examens*, París, Bourrellet, 1962.
- Ingenkamp, K., *Die Fragwürdigkeit der Zensurengebung*, Weinheim, Beltz, 1971.
- Laugier, H. y Schreider, E., "Recherche docimologique sur un examen de l'enseignement supérieur", en *Biotypologie*, 1958, 19, n° 2, 61-72.
- Lauwerys, J. A. y Scanlon, D. G., Ed., *Examinations*, The World Year Book of Education, 1969, Londres, Evans, 1969.
- Lloyd, W. A., "Les examens en Angleterre", en *Revue Française de Pédagogie*, en. 1968.
- Mather, D., France, N. y Sare, G., *The C.S.E., A Handbook for Moderators*, Londres, Collins, 1965.
- McIntosh, D., Walker, D. y McKay, D., *The Scaling of Teachers' Marks and Estimates*, Edimburgo, Oliver y Boyd, 1962, 2ª. ed.
- Montgomery, R. J., *Examinations. An Account of their Evolution as Administrative Devices in England*, Londres, Longmans, 1965.
- Joint Matriculation Board, *The Marking of Scripts in Advanced Level*

- History*, Universidades de Manchester, Liverpool, Leeds, Sheffield y Birmingham, 1964.
- Otter, H. S., *A Functional Language Examination*, Oxford Univ. Press, 1968.
- Passeron, J. C., "Sociologie des examens", en *Education et Gestion*, 1970, 2, 6-16.
- Pedley, F. H., *A Parent's Guide to Examinations*, Oxford, Pergamon Press, 1964.
- Pidgeon, D. y Yates, A., *An introduction to Educational Measurement*, Londres, Routledge y Kegan Paul, 1968.
- Pieron, H., *Examens et docimologie*, París, P.U.F., 1963.
- Piobetta, J. B., *Examens et Concours*, París, P.U.F., 1943.
- Remondino, C., "Recherche sur les systèmes numériques d'évaluation scolaire", en *Le Travail humain*, 1965, 18, 3-4, 263-265.
- Reports of the Secondary School Examinations Council*, Londres, H.M.S.O., 1947, (1st) - 1964 (8st).
- Reuchlin, M., *L'orientation pendant la période scolaire*, Estrasburgo, Consejo de Europa, C.C.C., 1964.
- Reuchlin, M. y Bacher, F., "L'appréciation des élèves par leurs professeurs", en *Revue française de Pédagogie*, 1968, 2, 19-25.
- Roller, S., "L'évaluation du travail pédagogique", en *Educateur et bulletin corporatif*, Montreux, 1970, 36, 694-696.
- Rot, N. y Bujas, Z., "Les distributions de notes scolaires comparées aux distributions des résultats obtenus dans les tests de connaissances", en *Le Travail humain*, 1959, 22, 19-26.
- Thorndike, R. L., "Marks and Marking Systems", en R.L. Ebel, *Encyclopedia of Educational Research*, Londres, McMillan, 1969, págs. 759-766.
- "Symposium sur la docimologie, XIII^e congrès de l'Association Internationale de Psychologie", Roma, 1958, en *Le Travail humain*, XXII, 1-2, en-jun. 1959.
- Tyler, R., Gagne, R. y Scriven, N., *Perspectives of Curriculum Evaluation*, AERA Monograph series on curriculum eval., N^o 1, Chicago, Rand Mc Nally, 1967.
- Valentine, C. W., *The Reliability of Examinations*, University of London Press, 1932.
- Vernon, P., *Secondary School Selection*, Londres, Methuen, 1957.
- Walker, A. S., *Pupils' School Records*, Newnes, Educ. Publ., 1955.
- Wendler, J., *Standard arbeiten, Verfahren sur Objektivierung der Notengebung*, Weinheim, J. Beltz, 1969.
- Wiseman, S., "The Marking of English Composition in Grammar School Selection", en *British Journal of Educational Psychology*, XIX, 1949, 200-209.
- Wiseman, S., *Examinations and English Education*, Manchester, University Press, 1961.
- Wrigley, J., "The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools", en *British Journal of Educational Psychology*, 25, 1955, 107-116.
- Yates, A. y Pidgeon, D., *Admission to Grammar Schools*, Londres, N.F.E.R., 1957.

INDICE ANALITICO

INTRODUCCION VII.

PRIMERA PARTE

DEFINICIONES

I. Docimología, docimástica y doxología	3
II. Exámenes y concursos	4
Observación y evaluación continuas	4
Exámenes internos y exámenes externos	6
III. Notas y <i>scores</i>	6

SEGUNDA PARTE

ACUSACION Y DEFENSA

CAPITULO 1 — Crítica de los exámenes	11
1. Cuerpos extraños en la educación al servicio de una pedagogía anticuada	11
2. Ansiedad y <i>stress</i>	12
3. Desigualdad — injusticia	13
4. El fracaso, generador de fracasos	18
5. Ruptura entre enseñanza y examen	20
6. Desacuerdo entre las personas encargadas de corregir la tarea escolar	21

a) La redacción francesa	21
b) Matemáticas	22
c) Medicina	23
d) Varios	23
e) ¿Cuánta gente encargada de la corrección haría falta para estabilizar la nota?	24
7. Inconsecuencia de una misma persona encargada de la corrección	25
Un esquema para continuar la investigación	26
8. Estereotipos y efectos de halo	27
a) Estereotipia	27
b) Efecto de halo	29
9. Falta de validez	31
10. Un instrumento de inmovilidad social	32
a) Efectos irreversibles del certificado escolar	32
b) Los exámenes no son socialmente neutros	33
11. Debilidad de numerosas experiencias docimológicas	35
12. Otras críticas	36
CAPITULO 2 — Defensa de la nota subjetiva y del examen	38
1. La medición rigurosa tal vez resulte imposible	39
2. Los maestros juzgan bien a sus alumnos	41
3. Validez limitada pero real de los exámenes tradicionales	43
4. Endurecerse para la vida	43
5. Ubicarse en relación con los demás	44
6. Amplia síntesis e integración de los conocimientos	44
7. El examen externo controla al profesor	44
8. El examen externo, <i>feedback</i> para el profesor	45

TERCERA PARTE

LA CONSTRUCCION DEL EXAMEN

Las grandes etapas. Vista de conjunto	48
CAPITULO 1 — El objeto y los objetivos	49
I. El objeto	49
A. El pronóstico	50
1. <i>Tests</i> de madurez específica (<i>readiness</i>)	50

2. Verificación de conocimientos clave o de nociones críticas	51
3. Ensayo	51
B. El inventario	51
C. El diagnóstico	52
II. Los objetivos	53
A. Los objetivos generales	55
1. Los objetivos cognoscitivos	56
a) La taxonomía de Bloom	56
b) El modelo de J. P. Guilford	59
2. El dominio afectivo	63
B. Los objetivos especiales	65
Ejemplos	67
 CAPITULO 2 — Redacción de las preguntas	 73
I. Observaciones generales	73
A. Preguntas comprensibles	74
B. Tener en cuenta el nivel de información	74
C. Poner a prueba o "testear" las preguntas	75
D. Cálculo de la facilidad de las preguntas	75
E. Cálculo de la eficacia. Poder discriminatorio	75
1. Método simple	76
2. Método más sutil	76
II. ¿Respuestas abiertas o cerradas?	80
A. Respuestas abiertas	80
B. Respuestas cerradas. Preguntas de elección múltiple	82
1. Utilidad	82
2. Elaborar una lista de preguntas	82
3. Explotar la gama de posibilidades lógicas	83
a) Pregunta con complemento simple	83
b) Asociación simple	84
c) Asociación compuesta	84
d) Asociación con término excluido	85
e) Análisis de la relación de causa a efecto	85
f) Análisis de observación	86
g) Comparaciones cuantitativas	88

h) Relaciones	88
i) Complementos agrupados	88
4. Cálculo de la eficacia de los "distractores"	89
a) Situación ideal	89
b) Situación por corregir	89
5. Críticas y refutación parcial	89
a) Una objetividad engañosa	89
b) Elecciones "correctas" controvertibles	90
c) Un juego de azar	91
d) Acrobacia mental	92
e) Inconvenientes inciertos	93
C. A manera de conclusión: un compromiso	93
 CAPITULO 3 — La calificación	 95
I. Un preámbulo indispensable: la curva de Gauss	95
A. La curva de Gauss, imagen de la probabilidad	95
B. La curva de Gauss, imagen de los resultados de la enseñanza no individualizada	97
C. El desvío tipo (estándar) o sigma, índice precioso	98
1. Significado	98
2. Estimación rápida de la medida y del sigma	99
a) Problema	99
b) Método sencillo de cálculo	101
D. La concentración de los resultados alrededor de la media	102
E. Curva de Gauss reclamada por los maestros	103
F. ¿Cómo se sabe si una distribución es normal?	103
II. La calificación subjetiva: escala de evaluación	107
A. Introducción	107
B. Indole y debilidades de las escalas de evaluación	108
C. Utilidad	110
D. Construcción	110
1. ¿Cuántos grados?	110
2. Definir el objeto de evaluación	111
E. Utilización	114
1. ¿Cuántos alumnos por escalón?	114
— El alumno comparado consigo mismo	114
— Alumnos comparados entre sí	116
2. Luchar contra la contaminación y la tendencia central	117
F. ¿Cómo sintetizar las evaluaciones?	118

III. L
IV. E

A
B
C
D

CAPITULO

I
2
8

CAPITULO

I. I
II. I

CAPITULO

G. Un caso particular: la calificación de la redacción francesa	121
1. Cuatro métodos de evaluación	122
a) El método de la impresión general	122
b) Las escalas de modelos	124
c) El método analítico	125
¿Qué cualidades se observarán?	126
Un ejemplo	128
d) El método de los recuentos de frecuencias	131
2. ¿Varios temas para elegir?	132
3. Conclusión	134
III. La calificación objetiva	135
IV. Estandarización o medida de la posición relativa	135
A. Los percentiles	136
B. Las notas estándar o notas Z	139
C. La escala de cinco clases	142
D. La escala de nueve clases (<i>stanines</i>)	142
 CAPITULO 4 — Control de la fidelidad del examen	 144
1. Evitar toda ambigüedad en las preguntas	144
2. Un número suficiente de preguntas	145
3. Un control matemático	145
a) El método pares-impares	145
b) Dos formas paralelas	146
4. Repetición de la calificación	146
 CAPITULO 5 — Control de la validez	 147
I. La validez del contenido	147
II. La validez de predicción	150

CUARTA PARTE

LOS PROCEDIMIENTOS DE MODERACION

CAPITULO 1 — Planteo del problema	155
1. Definición	155
2. Moderar no es someter a disciplina rígida	157
3. ¿Moderación espontánea o impuesta?	157
4. La moderación empieza junto con el año lectivo	157
5. No existe comparación posible sin elevada fidelidad	159
6. ¿Podemos confiar en los <i>tests</i> ?	159

CAPITULO 2 — Algunos sistemas de moderación de exámenes	161	
I. La fórmula más liberal: el sistema sueco	161	
II. Sistema de moderación establecido a partir de un <i>test</i> de conocimientos	164	II.
III. Moderación espontánea mediante apelación a un “banco” de preguntas	165	
IV. Un sistema sencillo de moderación global al servicio de la selección	166	
V. En Inglaterra, un sistema completo de moderación	169	
A. Preliminares	169	
B. Los profesores califican sus exámenes	170	III.
C. Nueva corrección de una muestra de trabajos por parte de los moderadores	170	
1. Control de la severidad	174	
2. Control de la discriminación	175	
3. Control de la conformidad	175	
D. Nueva corrección de las muestras restantes y control	175	
E. ¿Cómo ajustar las notas discordantes?	178	
1. Ajuste de la mediana	178	
2. Ajuste de la mediana y del desvío tipo	178	
VI. La nota de fin de año (Trabajo del año + trabajos prácticos + <i>test</i>)	181	

QUINTA PARTE

EL MITO DE LA CURVA DE GAUSS

CAPITULO 1 — Evolución de la curva de conocimientos	190
I. La curva de aptitudes	190
II. La curva de conocimientos	191
CAPITULO 2 — Una pedagogía de la curva en J	194
CAPITULO 3 — La teoría de la evaluación formativa	197
I. Jalonar la ascensión al saber	199
II. Guiar al alumno	200
III. Límite práctico del sistema	201
IV. El sistema de enseñanza puesto en tela de juicio	201
CONCLUSIONES Y RECOMENDACIONES	203

APENDICES

I. Estudio comparado de una pregunta de examen presentada según el método tradicional y el de elección múltiple	211
---	-----

— Método tradicional	211
— Método de preguntas de elección múltiple	212
II. Redacción en lengua materna — Inglaterra, G.C.E., 1967 — Fin del ciclo secundario — Nivel común	217
— Preguntas	217
Prueba I	217
Prueba II	218
— Consignas para la corrección	220
Prueba I	220
Prueba II	222
III. Ejemplo de enseñanza semiindividualizada	226
— La New Trier Township High School	226
— La individualización de los programas	227
— La individualización de la enseñanza	230
— El estudio de cada alumno antes de su ingreso en la New Trier	236
— Los consejeros pedagógicos	238

Impreso en el mes de diciembre de 1973
en Lauka Impresiones, Colombres 1574,
Buenos Aires. Edición de 3000 ejemplares.

GILBERT DE LANDSHEERE

EVALUACION CONTINUA Y EXAMENES
Manual de docimología: un nuevo desarrollo
conceptual y práctico de ambos procesos

Reaccionando contra las corrientes tradicionales, el autor se preocupa por elaborar una evaluación constructiva, positiva. Por primera vez, los docentes a quienes esta obra está dirigida dispondrán de criterios simples para construir los instrumentos de evaluación continua y para sintetizar e interpretar los resultados. Esto permite hacer comparables las notas de diferentes maestros y de distintas escuelas.

El libro es útil para los docentes de todos los niveles, que podrán poner en práctica su contenido sin ningún conocimiento especializado de estadística.

“La época del antagonismo entre psicómetras y educadores ha tocado a su fin —afirma el autor, profesor de la Universidad de Lieja—. Los primeros deben humanizar sus números; los segundos, introducir más rigor en sus procedimientos; los dos han de unir y poner en armonía sus esfuerzos, en beneficio del estudiante y de la comunidad.”

