

PROBLEMAS PSICOLOGICOS DE LA VALORACION DEL RENDIMIENTO ESCOLAR

Por DR. KLAUS HASEMANN *

1. *Advertencias generales.*

El problema de la valoración del rendimiento escolar es tan antiguo como la escuela misma. Este problema existe en todos los niveles de la enseñanza; pertenece así a la escuela primaria como a la universidad, y aparece siempre allí, donde es necesario describir el valor de un rendimiento, que es efectuado por medio de los que toman parte en cualquier enseñanza. Por esta razón el problema siempre ocupa un lugar preferente entre los temas de las revistas especializadas y libros pedagógicos y psicopedagógicos.

Lamentablemente los diferentes métodos de calificar fueron y siguen siendo objeto de disputas filosóficas e ideológicas. Esto ha dificultado que uno se haya ocupado objetivamente en este problema aumentando así la inseguridad sobre la validez de los sistemas de notas.

Es evidente, a pesar del intenso interés que existe sobre el problema, que éste no ha sido solucionado. Cuando se revisan publicaciones sobre el tema "calificaciones y notas" de los últimos cincuenta años, se trata siempre, una y otra vez de las mismas preguntas. Estas preguntas se pueden diferenciar en algunos aspectos. Unos cuantos de estos se pueden formular de la siguiente manera (véase Ebel 1965 y otros):

a) ¿Qué se entiende por "trabajo o rendimiento escolar"?

b) ¿Tienen las notas la finalidad de señalar solamente el valor del trabajo o también las actitudes, propiedades y otros rasgos del carácter?

c) ¿Tienen las notas el objeto de mostrar la evolución y desarrollo del rendimiento o su estado en un cierto momento?

d) ¿Es oportuno valorar diferentes aspectos del rendimiento con una sola nota o cada aspecto separadamente?

e) ¿Las notas deben darse conforme a una medida absoluta, en la cual la dimensión superior es la maestría, y en otro extremo está el fallo? O es mejor tomar como medida los rendimientos, que grupos de personas han absuelto y ¿qué procedimientos se usan para comprobarlos?

f) ¿Debe tener la escala de notas pocos o muchos grados?

No me será posible por la brevedad del tiempo que dispongo, ventilar todos los aspectos del problema arriba mencionados. Por este motivo permítanme elegir aquellos que son especialmente rele-

* *Del Abteilung für Angewandte Psychologie, Psychologisches Institut der Universität Freiburg (Breisgau), Alemania.*

vantes para el punto de vista psicométrico, el que está en el centro de mis reflexiones.

Rendimiento escolar es (según Just 1934, pág. 166) la manifestación de una cierta disposición intelectual, del carácter y corporal, que resulta bajo determinadas condiciones de la evolución. Esta definición del concepto es una de biología hereditaria y teoría ambiental, y ella es suplida por la definición operacional, que interpreta rendimientos escolares como resultados de un experimento continuado, el cual se extiende durante mucho tiempo y el cual transcurre bajo un control permanente; de este modo la ejecución está garantizada hasta una cierta parte, por medio de reglamentos (Just 1939, pág. 539). En este sentido, a saber, como resultado del trabajo del alumno, estudiante o aprendiz, es decir, como resultado de una conducta, se debe comprender el rendimiento en el presente trabajo.

Teóricamente se puede distinguir entre la verificación del rendimiento individual, la valoración y la comunicación. Por ejemplo, es imaginable que el modo de juzgar es un contar de faltas o una consideración en total. O también puede ser que el profesor tome la valoración de un rendimiento particular, de diferente manera a lo que la comunicación da a entender. Casi siempre la valoración dada es el producto de una serie de valoraciones particulares.

La forma más frecuente de la comunicación en el juzgamiento de las pruebas sobre el trabajo y la conducta del alumno es el certificado.

Sobre el valor, el significado y la eficacia declarativa del certificado, y sobre las censuras en el contenido, han discutido mucho los pedagogos, psicólogos y sociólogos.

Para el resultado de todas estas discusiones es casi imposible realizar una concordancia en la valoración de rendimientos. Esto vale particularmente para rendimientos iguales aun en idénticos establecimientos de enseñanza con todos

los alumnos. También si se define claramente el objeto a juzgar y se explica cada grado de las notas lo más exactamente posible, no se llegará a una comparabilidad.

Para conseguir esta comparabilidad sería necesario eliminar varios obstáculos. Uno de estos obstáculos son las diferencias individuales, que hay entre cada curso, que existen en la estructura de la población y en las exigencias del currículum. Así también se tiene que nombrar las diferentes pretensiones de los profesores como sus conceptos sobre la escuela y la enseñanza. Otras desigualdades resultan de las diferentes relaciones que hay entre el profesor y su curso, como también por los desvíos individuales en los principios de la valoración. ¿En dónde se tiene en cuenta realmente la conducta inicial individual, el modo de influir pedagógicamente y los efectos sobre el proceso de aprender? ¿No es más bien así, como algunos autores han visto a base de experiencia e investigaciones, que las censuras no son en primer lugar un juicio sobre el trabajo del alumno, sino una expresión de la satisfacción del profesor con el alumno (Gaupp 1963)? ¿No se censura más bien la reacción del alumno a la personalidad del profesor, al modo pedagógico con el que el profesor dirige al alumno, al modo metódico de transmitir la materia? Cuando se reflexiona sobre todo esto, parece improbable, que aquella crítica unida en notas pueda contener un fondo objetivo (Sander 1963, pág. 475).

En cuanto la objetividad de la valoración se fija como medida de concordancia entre varios críticos, la clasificación de los mismos rendimientos, no se puede esperar en base de resultados de investigaciones contruídos correspondientemente, ni se puede esperar que en la nota dada por el profesor se tenga una segura declaración sobre el valor del trabajo o examen del alumno (véase Blumenfeld 1931; Hartog y Rhodes 1936; Vicenci 1953; Kapuste y Sund 1966; Reuchlin 1967).

El problema central es aquí, poner el juicio del profesor, que aparece en la nota, en forma objetiva. Se trata con todo esto de la posibilidad de comparar las valoraciones de "prácticas adquiridas" y más precisamente, de la posibilidad de comparar pruebas de trabajo y conducta sobre las cuales se pueda deducir una aptitud.

Wagenschein (1954, pág. 411) define la nota de la siguiente manera: "Una nota es el resumen de un juicio escolar en un número". El punto o zona sobre la escala, que la cifra designa, se caracteriza generalmente por una denominación, como por ejemplo "siete" igual "excelente", "seis" igual "muy bien" etc. Pero esta denominación tiene solamente valor nominal, ya que no se halla en el caso de caracterizar sin equivocación la sección que a ella le pertenece. Unas de las causas principales para la imperfección en casi todos los sistemas de valoración del trabajo escolar es que los grados singulares de la escala censural son por regla general raramente bien definidos. Para esto pueden hacerse demostraciones empíricas.

Por ejemplo Travers y Gronlund (1950) encontraron en profesores universitarios grandes diferencias entre los conceptos sobre la importancia de unas notas como también sobre las reglas para juzgar.

Ebel (1965) informa de asombrosas diferencias en la distribución de las notas en diferentes departamentos de colleges.

Los diversos motivos para la inseguridad sobre el significado de los predicados en palabras, han sido muchas veces discutidos críticamente. Meese (1922) advierte sobre la inconsecuencia de algunas formas de marcar grados en notas: Si existe una nota "suficiente" pero no un "satisfactorio", entonces tampoco se debiera nombrar la peor nota "nulo" sino "insuficiente". Es imaginable, que un rendimiento es satisfactorio al destino del curso y se la denomina como "muy bien", porque la palabra "satisfactorio"

contiene en el uso del idioma por cierto un juicio positivo; así también un buen rendimiento es pues "satisfactorio", pero también puede tener algunas deficiencias. Por lo tanto la praxis de censurar quiere caracterizar con esta designación un rendimiento ni positivo ni negativo, y por consiguiente una nota en el medio exacto de la escala. Cuando se usan aquellas palabras para valoraciones, siempre tendrá que ser uso inequívoco, y señalar que un rendimiento es todavía suficiente o ya no satisfactorio (Huth 1955).

Las opiniones sobre la utilidad de usar cifras para la valoración de rendimientos discrepan. A un lado están aquellos —como Pestalozzi— que rehusan la calificación en cifras por motivos pedagógicos (por ejemplo Schreiber 1899; Zeidler 1928; Döring 1931; Kleinert 1951; Merz 1951; Simoneit 1952; Ruppert, 1954; Zillig 1959; Bernart 1962; Flitner 1966).

Al otro lado domina la idea, de que las cifras tienen varias ventajas, por lo que se las usa con preferencia en la praxis escolar y esto con razón. La cifra es fácil de manejar, muestra sin duda un punto o una zona sobre la escala; es una expresión unívoca, con la que se puede comparar sin dificultad los rendimientos, y permite además una comprensión fácil e independiente del lenguaje por terceras personas (Kiessling 1929; Reiring 1954; Körner 1956; Engelmayr 1960; Diedrich 1961; Dohse 1963; Göller 1966). Estas preferencias hacen las notas, en cifras, utilizables como datos para investigaciones científicas.

Se deja objetar críticamente contra esto, por ejemplo, que los puntos de vista y las reglas para el modo de censurar no se pueden inferir del término numeral, ni tampoco que aspectos de la materia respectiva en la valoración estaban en primer plano. Sin embargo se puede dar razón a Wagenschein (1954, pág. 412), cuando hace constar, que la nota en cifras sea "indispensable en la rápida selección, despacho, aprecio y dis-

tribución de grandes masas de aspirantes, que hoy es inevitable. Por consiguiente es económico. ¿A dónde iríamos a parar, si cada vez tuviéramos que leer un pequeño dictamen!”.

Hay siempre el peligro, de que juzgamientos verbales sean mal entendidos por el destinatario. De ordinario se quedan en un angosto margen de lugares comunes... en aglomeraciones de palabras preponderantemente feas... que siempre se repiten” (Hermann 1948, pág. 738).

Solo el hecho de que el sentido de una palabra es ambiguo puede dar fácilmente lugar a controversias sobre el grado de una cualidad. Al contrario descripciones y características del decurso y resultado del trabajo escolar como también comentarios a las notas en cifras pueden facilitar el absoluto entendimiento de una valoración.

Otra ventaja al usar cifras para certificar es su función separativa en las diferentes zonas de rendimiento. Esta función permite concretar, justamente para la selección, los rendimientos mínimos, los cuales tienen que llegar a ser logrados por ciertos grados de habilidad.

Ya que la valoración de rendimientos siempre se realiza por comparación, ocupa la comparabilidad el primer lugar en la calificación.

Verdad es que la comparabilidad entonces, tampoco debería ser reducida por diferencias en las reglas de la valoración, referidas a las regiones, ramos o personas. Esto vale especialmente para el certificado de bachiller como documento de la habilidad para un estudio universitario.

2. *Problemas sobre la escala en la valoración del rendimiento escolar.*

Las discusiones sobre la manera de dividir mejor las escalas de la valoración y evaluación se sucedieron en la psicología como también en la pedagogía al principio de la segunda década de este siglo.

En los rendimientos escolares se pregunta desde entonces por la cantidad óptima de grados sobre la escala de notas. Blumenfeld (1931) ha demostrado, que la división de la escala de notas en los dos puntos de inflexión de la curva de Gauss, es matemáticamente fundada. Empezando por esta “forma original” de la escala de notas (“promedio”, “debajo del promedio”, “sobre el promedio”) han sido propuestas hasta ahora casi todas las posibilidades de la división de la escala, y estas fueron practicadas en un rápido cambio, generalmente para desventaja del alumno. En general se confrontan conceptos contrarios. Por una parte se ve como óptimo un pequeño número de grados, como por ejemplo Vincenci (1953) recomienda la limitación a tres grados; por otra parte se rechazan las escalas con cuatro a seis grados por no ser muy diferenciables. Advirtiendo en la numerosa cantidad de graduaciones posibles en los rendimientos se ven diez, quince o más puntos como la adecuada división (por ejemplo de Schott 1920, 1922; Franke 1922; Sost 1926; Strohmeier 1930; Merz 1951; Körner 1956). A principios de este siglo eran usuales en los Estados Unidos por ejemplo escalas de notas que teóricamente permitían emplear todos los valores enteros entre un rendimiento cero por ciento y uno de cien por ciento. Pero prácticamente se utilizaron solamente treinta de estos cien puntos posibles. Según investigaciones de Conklin (1923), Symonds (1924), Champney y Marshall (1939), Hake y Garner (1951) y otros es el óptimo para escalas de valoración o sea evaluación entre cinco y nueve grados.

De la práctica de dividir las notas (por ejemplo interponer entre la nota “dos” y la nota “tres” una nota “dos coma cinco”), nació a menudo, de una escala con cinco grados, una de nueve o diez grados. El traspaso del sistema de notas al sistema de puntos es pues corriente. Para la valoración de rendimientos singulares puede ser útil graduar la

escala de cero hasta diez o hasta veinte o, quizás, hasta cien puntos. Este es el caso, si se tiene que ver con tareas, cuya solución se compone de varias partes particularmente valorables. Tampoco se debe desatender, que una reducción de la cantidad de grados puede aumentar las faltas sistemáticas de la valoración. Es más injusto dar una nota "cuatro" de una escala con cinco grados a un alumno, que en realidad, —es decir objetivamente—, debería tener "tres", que dar un "ocho" de una escala con veinte grados a un alumno, que objetivamente hubiera merecido un "diez". Al fin y al cabo las circunstancias reales se dejan probablemente ilustrar mejor por medio de sistemas de notas con muchos grados, pero la valoración por puntos encierra riesgo de una diferenciación aparente (véase Ruch 1929), porque el problema de la definición de cada grado aparte de la escala, debería ser insoluble para escalas de múltiples grados.

Para la construcción de escalas de notas valen igualmente las mismas reglas como para las escalas de evaluación psicológica, en inglés llamadas "rating scales". Esto quiere decir, que la definición de acercarse a características ilustrables del rendimiento, debería ser obtenida por extensas investigaciones previas. En este proceso se tienen que elegir con esmero las más útiles de numerosas definiciones. Recién la interpretación estadística de estas investigaciones previas dan resultado, si una definición es o no precisa. Solo definiciones unívocas, que se separan exactamente y que no necesitan explicaciones suplementarias permiten un uso más o menos concordativo.

Por lo menos tienen que ser definidos los rendimientos en cada decimal de la escala como también el sentido de cada paso de punto a punto. Pero casi nunca se podrán encontrar tantas diferentes designaciones, que separen cada grado claramente, como son necesarias para una escala con tantos grados. Pero cuando no se define cada grado de la escala,

entonces se usan menos los grados no definidos.

De esto resulta un cambio falso de sitio en la distribución de las notas. En especial para el bachillerato, la aplicación de un sistema así es muy dudoso, porque sobre rendimientos tan complejos como el bachillerato, exige, la aptitud y la buena voluntad de los críticos; en este caso: los profesores, son bastante limitados para dar tales críticas diferenciadas (véase Guilford 1954). Un sistema de notas con pocos grados sobre la escala fue al fin usado tanto en las regiones anglo-americanas (cinco letras de A hasta F) como en las alemanas (cifras uno hasta cinco o seis). Mientras algunos autores dudan si es apropiada la división de la escala de un número impar en un número par, otros autores ven en esta diferenciación, por lo menos contra la escala de cinco grados, una ventaja más grande.

La falta de una nota media —de todas aquellas críticas de escalas pares vistas como desventajadas— puede ser denominado justamente como una preferencia de este tipo de escala, cuando se toma en consideración, también en la calificación escolar, la influencia de los más frecuentes errores sistemáticos de la valoración y evaluación (véase Guilford 1954; Lienert y Hopp 1964). Por ejemplo si les quita a los juzgadores la posibilidad de seguir su tendencia a dar censuras "neutrales", como son las del medio (véase Hasemann 1964). Cuando hay solo cinco notas sobre la escala, entonces se tiene que valorar con la nota "tres", tanto los rendimientos que tienden al promedio un poco hacia los buenos como los que están un poco debajo del promedio. Otro error en las valoraciones, que a menudo se presentan, es el "error de la lenidad", en inglés llamado "error of leniency", es decir la calificación demasiado buena. Este error es, entre otros, el motivo por el cual el porcentaje de alumnos con notas sobre el promedio, es desproporcionadamente grande en muchos exámenes, especial-

mente en exámenes al finalizar un estudio (véase Sost 1926; Bessler 1927; Lämmermann 1927; Blumenfeld 1931; Strunz 1967).

El hecho de que la “tendencia al centro” se deje eliminar por el uso de escalas con grados pares, es ciertamente una ventaja. Sin embargo los de opinión adversa tienen razón al decir, que se da un poco al capricho —suponiendo distribución “normal” sobre las seis notas de la escala— la nota “tres” o “cuatro” a un cierto porcentaje de los alumnos (cerca de veinte por ciento), el cual está muy unido alrededor del promedio efectivo. Conforme a esto, cada quinto alumno sería asociado a la mejor o a la peor mitad del curso, por falta de una nota en el centro de la escala. Por esto se explica en parte, que la distribución es diferente de profesor a profesor. Porque la “tendencia al centro” suele ser fuerte, casi todos los juzgadores —en el caso de notas escolares: los profesores— buscan una censura utilizable como nota central. Entonces tienen que decidirse por un “cuatro” o un “tres”. Verdad es que esta manera de ver el problema es métricamente exacto, pero ello no es confirmado por el uso real del modo de dar notas.

Hopp y Lienert (1965) y también yo en mis investigaciones del año pasado (Hasemann 1968) pudimos demostrar, que la nota “seis” nunca es dada; en algunas disciplinas aun la nota “cinco” tenía un porcentaje casi nulo (las notas seis y cinco son las notas más bajas en Alemania). Estos resultados se refieren a los cursos superiores de la secundaria, y ello da motivo para hacer constar: el sistema que, sobre el papel, tiene seis grados, en la práctica tiene solamente cinco grados (véase también Bartel 1961). La limitación que los profesores ejecutan evidentemente por sí solos, tiene su motivo, en algunas disciplinas, probablemente en el carácter del objeto de las lecciones —por ejemplo, en religión, dibujo, música, gimnasia—, en las cuales la tendencia de dar notas solo en tres

grados (“mal”, “medio”, “bien”) es comprensible.

Si investigamos los motivos, que hacen difícil resolver el problema de la valoración escolar, tenemos que considerar el hecho, de que se trata de las censuras por resultados de mediciones. Por una parte los pedagogos rehusan la aplicación del término “medir” al proceso de valorar el trabajo del alumno. En la práctica educativa se une a menudo con el término “medir” la impresión de una esquematización. Por otra parte los psicólogos son escépticos con el método escolar de concretar los grados de rendimiento y con sus resultados. Cuando uno acepta la teoría mensural de Stevens (1951), tal vez la más aprobada en la psicometría, entonces se tiene que hacer constar como psicólogo, que el proceso censural puede ser visto del todo como medición en un sentido amplio. Medición es, según Stevens, una “agregación de números a objetos o acontecimientos en concordancia con reglas” (pág. 1). A esta definición general se adjunta la pregunta, ¿en qué nivel tiene lugar la medición? Es decir en el caso de censuras, ¿si en principio, bastan las notas a las exigencias que hoy se dan a números mensurales para evaluar rendimientos de personas?

De las causas para las dificultades en la medición de rendimientos psíquicos del hombre, Thorndike menciona en su libro “*Mental and Social Measurements*” ya en el año de mil novecientos doce las siguientes, como las más importantes: Primero, la falta o la imperfección de medidas, en las cuales se mensura; segundo, la falta de constancia en los objetos que tienen que ser medidos; y tercero, la complejidad de estos objetos.

Martinak había advertido en mil novecientos, que la escala de notas está dividida en “zonas no exactamente limitadas”, porque el punto en la fila de posibles grados de rendimiento no se dejan fijar de ningún modo exactamente, sino se deja solamente clasificar el rendimiento en zonas. La pregunta simple, si

las zonas tienen dimensiones iguales, no es contestable. La continuidad de los rendimientos está pues enfrente, en las notas escolares, a una escala discontinua. También Orlik (1961) y Pfanzagl (1964) han hecho constar, que es infundado el concepto usual de la escala de notas como escala mensural con distancias iguales (es decir como escala de intervalos). Orlik habla de la "incomparabilidad general" de las notas, con lo cual él entiende que el significado de las notas es diferente según las disciplinas particulares.

Finalmente para la escala de las notas vale decir, que un rendimiento que ha recibido la nota "dos" no es exactamente dos veces peor que un rendimiento certificado con la nota "cuatro". Tampoco se puede clasificar exactamente, si por ejemplo la distancia entre un rendimiento "muy bueno" y uno "bueno" es de igual dimensión como un rendimiento "deficiente" y un "mal".

Entonces, cuando se valoran rendimientos con ayuda de notas en cifras, se puede hablar solamente de un orden en rangos, es decir: "poner en grados por medio de una escala categorial, que es constituida ordinalmente" (Lienert y Hopp 1964, pág. 195).

El profesor usa esta escala para clasificar a todos los alumnos con rendimientos aproximadamente iguales en el mismo rango respectivamente. Este método solamente supone dada la posibilidad de poder clasificar rendimientos en rangos, o más bien la posibilidad de una diferenciación comparativa, es decir "mejor/peor".

3. *La medida de la valoración.*

Problemas parecidos como en la división de la escala, se plantean también aquí para el modo de la distribución, con el cual se distribuyen las notas efectivamente sobre los grados, que hay a disposición. Se tiene que preguntar por las normas conforme a las cuales se mide el rendimiento individual. Punto de pro-

cedencia es aquí nuevamente el fin del certificado, en el cual la nota es contenida: ¿Tiene el certificado el objeto de informar, si se logra o no se logra el propósito de la enseñanza? ¿Tiene el certificado la función de mostrar, qué lugar ocupa el rendimiento de un alumno entre los rendimientos de sus compañeros? ¿O debe cumplir las dos funciones a la vez?

Los dos puntos de vista más importantes, entre los cuales se puede ejecutar la distribución son por consiguiente: la medida teórica, es decir "absoluta", de exigencias fijadas y la medida empírica de rendimientos comprobados de un colectivo marcado.

Las dos posibilidades tienen sus opositores y sus defensores. Mientras algunos autores recomiendan una calificación que esté orientada al trabajo "distribuido normalmente" en un curso, otros autores recomiendan, que el rendimiento individual se debe medir no por medio de los rendimientos de los compañeros de curso, sino por medio del propósito de la enseñanza en el curso. Si se afirma el concepto arriba mencionado, entonces se tiene que convenir, en que no son suficientemente considerados los rasgos personales de cada alumno y el nivel diferente de los rendimientos de cada curso. Las experiencias nos enseñan, que hay cursos en los cuales los rendimientos tienen un buen nivel general o solamente en ciertas ramas, y que hay cursos con un nivel bajo; los dos casos pueden ser independientes del respectivo profesor. Sander (1963) pudo a través de una investigación con novecientos cincuenta alumnos de un colegio, hallar empíricamente seis diferentes tipos de rendimientos de curso.

"Medida absoluta" se llama a la necesidad de definir la cantidad de la materia en una rama, la cual debería ser aprendida en un cierto período. Se dice pues, que en el bachillerato deberá examinarse, si se logra o no se logra un cierto propósito de la enseñanza. Pero si esto se dice, entonces se podrá realizar

esta tarea solamente si está descrito el fin perseguido unívocamente. Este "propósito de la enseñanza" se deja hasta cierto punto hacer constante teóricamente, si se tiene un catálogo en el cual están fijadas las exigencias del examen en cada rama. Este catálogo debe poseer carácter obligatorio general y validez a plazo largo. Sin duda un catálogo de esta naturaleza puede ser válido como medida objetiva, no obstante su contenido debe ser necesariamente interpretado y por eso se queda en el arbitrio subjetivo. A causa de este hecho fracasan todos los sistemas de calificación, los cuales dan notas según una "medida absoluta". Tan difícil en limitar esta medida para arriba como para abajo, es decir responde a la pregunta, ¿cuánto tiene que haber aprendido una persona, para que ella responda apenas a las exigencias?

Las divergencias en las valoraciones de los mismos rendimientos son también motivadas por la diferencia que existe en el especial interés de la disciplina para los respectivos profesores. Diversos autores (por ejem. Vicenci 1953; Horney 1960) indican, que no hay coincidencia sobre lo que efectivamente se tiene que juzgar, lo que precisamente se trata en el rendimiento de cada rama.

Una norma objetiva del rendimiento por consiguiente no existe. Con respecto a la praxis de dar notas, también diferente de profesor a profesor, se tiene que mencionar fuera de los arriba dichos errores sistemáticos sobre la calificación, otro error en el juzgamiento, llamado "halo effect". Este efecto se nota como influencia en el rendimiento de un alumno en una disciplina a la valoración de su rendimiento en otra disciplina, o también como influencia de su inteligencia general y de su procedencia social a la valoración de su trabajo (véase por ejemplo Carter 1952).

Se hicieron varias y muy interesantes investigaciones para concretar el modo y la dimensión de esta influencia. Así pudo Sost (1926) por ejemplo verificar, que ochenta por ciento de los casos, juicios

sobre preocupaciones, talento y situación familiar entran en la valoración del rendimiento singular.

Zillig (1928) controló rigurosamente dictados de dos grupos de alumnos, unos juzgados en general como muy buenos alumnos y otros juzgados en general como muy malos. De las faltas de los buenos alumnos quedaron treinta y nueve por ciento inadvertidos, de los malos alumnos solamente doce por ciento, Weis (1965) dejó censurar composiciones y tareas matemáticas. El ofreció diferentes informaciones sobre familia, talento y costumbres de los alumnos a los profesores juzgadores de los mismos trabajos, que los alumnos escribieron. Las diferencias significativas, que en las calificaciones fueron dadas por los grupos de profesores separadamente informados, verificaron la influencia de factores extraños al rendimiento, sobre el juicio.

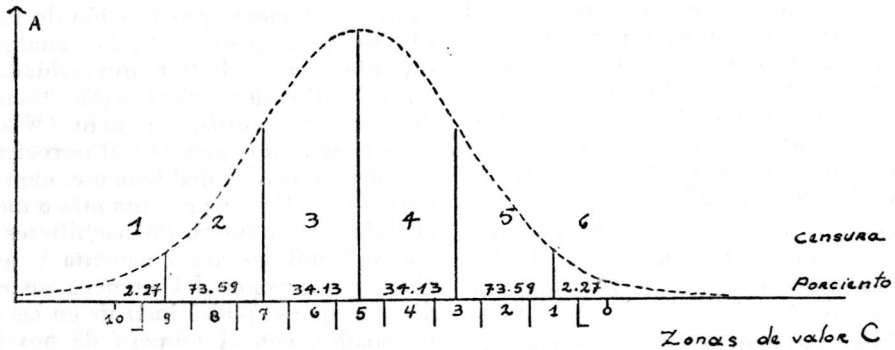
Además el diferente grado de complicación de las tareas impide el modo de dar notas, el cual está orientado a "exigencias", que no son empíricamente ganadas. La tasación del grado de la complicación de una misma tarea por medio de los profesores se ha mostrado como sumamente diferido (Lietzmann 1931; Bobertag 1933, 1934; Pieron 1963; Weiss 1965). Se hicieron propuestas con el fin de alcanzar sistemas de valoración, los cuales son convenientes al grado de complejidad en las tareas.

Para la realización de estas propuestas sería necesario mucho esfuerzo reformatorio y organizatorio. Para el bachillerato se ha probado disminuir un poco las influencias subjetivas por medio de una segunda valoración de la misma tarea o por relatos sobre la valoración de los rendimientos. Aunque justamente el llamado "bachillerato central" podría ser irresoluto desde el punto de vista pedagógico, son absolutamente necesarias aquellas o similares diligencias, con el fin de proporcionar una mensura de rendimientos estandarizados, para dar validez a las exigencias de una rama como un sistema referencial constante.

Con respecto a la medición de rendimientos en el trabajo y exámenes escolares sobre la medida empírica son imaginables diferentes modos de proceder. El problema de la fijación del grado de complejidad para cada tarea escolar es otra, que aquella para la distribución de la valoración de rendimientos según la "distribución normal", que para la valoración en vista a una intención absoluta de la enseñanza. Independiente del grado de complejidad de la tarea se queda en la "distribución normal" siempre igual a la frecuencia relativa con la cual un grado está ocupado (véase dibujo I). De esta manera se puede evitar que las valoraciones de rendimientos iguales en un mismo curso diverjan entre diferentes profesores. Solo la pregunta es: Si se mensura el grado del rendimiento individual por medio de relacionar este rendimiento con los rendimientos de los compañeros del mismo curso al resolver la

rización de tareas, como sería necesaria para una medición según normas. Por este motivo se podrá seguir por el momento solamente el camino sobre las comprobaciones de las relaciones internas del curso. Se deja precisar el lugar del porcentaje de cada rendimiento particular por medio de una serie de rangos, en los cuales los rendimientos dentro de un curso están distribuidos. Al lugar del porcentaje del rendimiento particular se puede dar entonces la censura respectiva.

También si se tiene en cuenta la "distribución normal" en efecto, el problema de la calificación escolar no es solucionado. ¿Es crítico preguntar, si se justifica dar la mejor nota de una escala a un alumno solamente, porque él ha obtenido el rendimiento relativamente mejor en un grupo de malos rendimientos? Igualmente necesaria como dar una va-



DIBUJO I: Frecuencia relativa de las censuras en la distribución normal.

A = Número de alumnos por cada censura.

C = Valor estandarizado de las censuras.

(Véase GUILFORD 1965).

misma tarea, o si se relaciona el rendimiento individual con los rendimientos de una gran muestra representativa en ciertos rasgos, ésta tomada como norma. En casi ningún caso de las pruebas del rendimiento como hoy generalmente se llevan a cabo (por ejemplo tareas en el curso, en la casa y tareas verbales) son dadas las condiciones para una estanda-

loración relativamente justa, sería saber en aquel caso, cuánto se ha acercado el alumno con su rendimiento al fin de la enseñanza. Una calificación, la que se basa sobre fundamentos empíricos como en la orientación de la "distribución normal", tiene la ventaja de que las relaciones dentro de la serie de las notas son sustraídas del arbitrio subjetivo. Pero no

se debe perder de vista, que la función de Gauss es la expresión matemática para la distribución de los casuales errores mensurales.

Pero para los cursos escolares, generalmente no se cumple la condición de la (bastante) gran cantidad, ni la combinación casual, como valen para la "distribución normal".

La subjetividad de la distribución de las censuras no se podrá evitar por completo. Ninguna de las modificaciones, que fueron recomendadas por varios autores (por ejemplo Höfler 1920; Bartel 1961; Weidig 1961; Göller 1966 et al.) puede satisfacer el postulado de objetividad, que siempre nuevamente se formula. La distribución de talentos y rendimientos es, pues, dentro de un curso, rara vez "normal" sino, por medio de diferentes causas pedagógicas, sociológicas y psicológicas, casi siempre desproporcionadas.

Al terminar estas nuestras reflexiones se tiene que mencionar, que ni por calificar por medio "de la distribución normal" ni por medio "de las exigencias del ramo", es decir por medio del fin de la enseñanza, se llegará a realizar exactos valores mensurales de rendimientos y constancia individual.

Parece, pues, que la deseada comparabilidad de los certificados que es absolutamente necesaria para la selección, por ejemplo desde el punto de vista de la habilidad para un estudio universitario, está fuera de alcanzarse. La distribución de las notas, como se presenta en su variedad y su irregularidad, tiene que ser evidentemente tomada como definitiva. La pregunta: "Hasta qué punto el certificado individual todavía permite una válida manifestación del rendimiento general de su propietario", es por consiguiente la pregunta por la posibilidad, siquiera de hacer mensurable algunas de las diferencias arriba mencionadas. Si se tiene una posibilidad así, entonces se podrá precisar aproximadamente exacta, la valencia de las notas en el certificado.

4. *Investigaciones para la clasificación científica de la valencia del certificado de bachiller.*

Existen de época reciente numerosas investigaciones empíricas de investigadores alemanes, que se ocupan con los problemas de la calificación y distribución de las notas (Pfaff 1961; Gaupp 1963; Sander 1963; Mock 1964; Weingardt 1964, 1967; Hopp y Liener 1965; Lucker 1965; Weiss 1965; Kapuste y Sund 1966; Orlik 1967; Spengler 1967; Schütz 1968). Pero solo pocos de ellos (Gaupp, Mock, Weingardt, Lucker, Orlik, Spengler, Schütz, Kapuste y Sund) se dedican con aquella valoración de rendimientos escolares, que será una de las más importantes, a saber, con la calificación como la que se presenta en el certificado de bachiller.

La importancia que tiene el certificado de bachiller por ejemplo en la República Federal Alemana, se caracteriza simplificada por función de legitimación para poder estudiar cualquier disciplina en cualquier universidad. De estos legitimados habrá según tasación del consejo científico alemán (Wissenschaftsrat), en el año de mil novecientos setenta y cinco probablemente, cien mil bachilleres. Esto representa más o menos el doble del número de bachilleres del año mil novecientos cincuenta y ocho. Una comparación del número de estudiantes, que se podrá instruir en las universidades, con el número de novicios, con el que tendrá que contarse en los años venideros por el crecimiento del número de bachilleres, muestra, que, también después de seguir con el desarrollo de las universidades, "las posibilidades de enseñanza no serán suficientes para la afluencia, que habrá en casi todas las disciplinas" (Wissenschaftsrat 1967, pág. 107).

Aquí existe una gran discrepancia. El bachillerato debe otorgar, tanto como antes, sin reserva, permiso a todas las disciplinas de la universidad. Las secundarias atestan efectivamente a los bachi-

lleser ser bastante "maduros" para dejar el colegio y estudiar. Pero de facto el certificado de bachiller no está definido suficientemente como documento de la "madurez para estudios universitarios". Como la cabida para estudiantes es limitada, las universidades se ven obligadas a seleccionar. Esta selección se tendrá que llevar a cabo "desde el punto de vista de la habilidad para la rama elegida" (Rektorenkonferenz 1964).

En el momento actual una selección que tiene en cuenta el punto de vista de la habilidad para la rama especial de un modo suficiente está lejos de ser ejecutada. Sin embargo parece ser así mismo necesaria una selección semejante, que se anuncie en la universidad, ya tiene el "permiso para el estudio universitario", a base de su certificado de bachiller propiamente dicho.

La medida más usual para la selección de los aspirantes, tanto con respecto al permiso para estudiar como con respecto a becas, son como antes las notas del certificado de bachiller. En la regla se da la preferencia al aspirante con el mejor certificado de bachiller. Diferencias en la cualidad del certificado, que resulta solamente por la diferente rigidez de la calificación, no se puede tomar en consideración, ya que al gremio, que ejecuta la selección, no le son conocidas estas diferencias. Pero, si también a base del certificado de bachiller, no se debe "calcular de antemano" hasta qué punto el bachiller satisface las exigencias de la rama que ha elegido, sin embargo no se debe menospreciar el valor de la función del certificado de bachiller como pronóstico en la situación presente.

Si la capacidad de las universidades fuera suficiente como para matricular sin excepción a todos los bachilleres que desean estudiar, entonces tendría que ocuparse intensiva y necesariamente de la pregunta: ¿Hasta qué punto las notas del certificado de bachiller son útiles como criterio de una evaluación, si los bachilleres cumplen con las condi-

ciones para un estudio universitario con respecto al rendimiento?

El certificado de bachiller debería bastar a los criterios de las mediciones psicodiagnósticas, para servir como fundamento de una evaluación de la habilidad. Para ser válido como medida objetiva de rendimientos, debería tener cada censura del bachillerato el mismo significado, es decir el mismo "valor", como medida evaluativa, no solamente en diferentes disciplinas, sino también en las mismas ramas con diferentes factores exteriores (por ejemplo, tipo escolar o región).

Como una parte para la solución del problema, si el certificado de bachiller puede cumplir con las condiciones arriba mencionadas, puede considerarse una investigación, la que yo realicé recientemente en la sección de Psicología Aplicada del Instituto Psicológico de la Universidad de Freiburg (Breisgau).

Para esta investigación se analizaron los certificados de bachiller de cada solicitante alemán, admitido en cualquier universidad alemana, el que hubo aspirado y fue matriculado en la Facultad de Filosofía de nuestra universidad entre el semestre de mil novecientos cincuenta y seis y el semestre de mil novecientos sesenta y cuatro. En total fueron explorados los datos de cuatro mil ciento sesenta y nueve bachilleres.

El criterio de la selección más usual para escoger solicitantes apropiados para estudios universitarios, o becas, es el promedio individual de algunas o todas las notas del certificado de bachiller. Orlik (1967) menciona sobre este hecho muy oportunamente: "Esta manera de proceder convence directamente, pero se empieza a poner en duda la legitimación de este procedimiento probablemente motivado por el método estadístico continuamente perfeccionado" (pág. 32).

Desde el punto de vista matemático Riebesell (1919 y Pfanzagl (1964) han indicado, que una adición de censuras para calcular promedios es irresoluta. Una contracción de notas singulares a

una resultante nota general, que responde a la realidad, no es posible a menos que los rendimientos valorados con las notas singulares, correlacionen muy bien. Una medida, la que es calculada por medio de promediación de valores así heterogéneos como son las notas escolares, es demasiado compleja como para ser fundamento para manifestaciones exactas (Blumenfeld 1931; Weiss 1964, 1965; Todt 1966).

Cuán indiferente sería para las decisiones, si se consideraran o no, las ramas particulares, fijadas en el certificado; entonces seguramente sería más económico, ya no ejecutar el examen y la calificación en estas disciplinas. Se tiene que dar razón a Orlik (1967) si él dice, "tales unidades funcionales complejas como la habilidad para estudios y la profesionalización no se dejan expresar en un sólo índice, pues justamente el promedio total del certificado de bachiller" (pág. 44). En cada motivo práctico, en el que se usa el certificado de bachiller, ciertas cualidades son de influencia preponderante. Una información sobre estas cualidades puede dar, a lo sumo, solamente la nota de un ramo o las notas singulares de una combinación de ramas. Por eso el concepto del certificado del bachiller como perfil del rendimiento promete resultados más inatacables, que una precisión de la valencia del certificado a base de censuras promediales (véase Strunz 1967).

La elaboración y la interpretación diferenciada de perfiles censurales son, sin embargo, solamente posibles, cuando se puede primeramente desatender el caso particular de este u otro colegio, de este u otro profesor, y en segundo lugar, tomar como base una gran cantidad de censuras, que diferentes profesores han dado a muchos alumnos (Sander 1963). Con ayuda de estadísticas representativas de esta clase se llega a una comparabilidad fundada empíricamente de las notas singulares de las ramas. Esta comparabilidad entre los tipos diferentes de escuelas, entre las diferentes regiones, si

es posible aun entre escuelas singulares de mismo tipo, facilita un análisis individual del perfil de rendimientos, como se documenta en el certificado de bachiller.

La comprobada distribución normal de la habilidad como la continuidad de las variables, la que se mide por motivos ya mencionados solamente en rudas clases, tienen la siguiente ventaja: ellas permiten la normalización estadística de valoraciones sobre rendimientos.

Esto es posible sobre la base de una prueba hecha representativa de la totalidad de rendimientos escolares temporales muy limitados. Los valores, que se reciben después de la normalización, permiten una comparación entre las disciplinas.

La calculación de valencias estandarizadas de notas requiere extensas estadísticas de certificados de bachilleres. Estas estadísticas deben ser separadas para cada combinación de categorías de las principales variables, sobre las que se fundan las diferencias en la calificación.

Nuestras investigaciones se basaron sobre estadísticas de esta manera. Sobre sus fundamentos, pudo ser realizada la substitución de los valores primordiales por medio de valencias estandarizadas. Después que nosotros hubimos ejecutado este cálculo, resultó realmente que las diferencias de calificaciones entre las ramas, tipos de escuela y regiones, se caracterizaron de esta manera de un modo más conciso que por promedios censurales. Pudo, pues, ser comprobada la valencia desigual de la calificación. Debido a este resultado, no se permite hacer decisiones para el certificado de bachiller a base de promedios individuales de valores primordiales de las censuras.

Los resultados de nuestra investigación ponen en duda la pretensión del certificado de bachiller, de ser documentación de la habilidad para un estudio universitario. No se justificó, pues, evaluar el certificado de bachiller como resultado de una medida objetiva del rendimiento. Esta "medida" no basta a

las condiciones entre las que es posible juzgar la habilidad de una persona. Confiabilidad, validez y objetividad de las notas en el certificado de bachiller no alcanzan el grado suficiente para preparar decisiones individuales e institucionales con respecto a la educación profesional en universidades.

Tampoco la estandarización de las notas puede dar ilusión sobre el hecho arriba mencionado. El postulado de que cada alumno talentoso, pero sólo éste, debería tener la autorización de ingresar a las universidades, se puede realizar únicamente con éxito satisfactorio, si es posible emplear métodos seguros para fijar talentos. Aquí se trata de abarcar conocimientos y comprensión, pero también de objetivar las posibilidades individuales, analizar problemas, mostrar métodos para resolver tareas y problemas, aplicando lo aprendido a diferentes circunstancias en pocas palabras: de la habilidad de pensar y trabajar científicamente.

Sin duda apenas es posible actualmente pensar en pruebas psicológicas de habilidades para la admisión en la universidad. Considerando que sería necesario examinar a todos los egresados de secundaria con tales pruebas psicológicas, y considerando las dificultades en la medición de la capacidad del rendimiento intelectual, en primer lugar se deberá desarrollar tests sobre el rendimiento escolar para el nivel de los cursos superiores de secundaria. Este desarrollo es tanto más importante cuanto más grande sea en los próximos años la parte de la enseñanza programada. La consecuencia evidente de esta nueva forma de enseñanza es el examen programado, que hace indispensable por consiguiente una

valoración objetiva (véase Wölker, 1968).

Tests de rendimientos escolares, para una selección de habilidad para pasar a la universidad existen en los Estados Unidos ya hace mucho tiempo (Seelig 1967), y se han aprobado desde hace diez años también en otros países, pero allí solo vacilantemente.

Pedraza y Baca (1965), Kohan (1968), Grassau et al. (1968) han comunicado recientemente sobre los trabajos correspondientes en países suramericanos. Estos tests se elaboran en los Estados Unidos por instituciones que son independientes del estado, de los colegios y universidades. Los tests se desarrollan según la forma más moderna de la construcción de tests psicológicos y se revisan permanentemente en su valor pronóstico. Para la medición actual, las tareas se componen siempre de nuevo. El resultado de la medición de rendimientos, que se ejecutó con los tests, es decisivo para la admisión o rechazo del aspirante por la universidad. La cantidad de puntos, que es determinante para la admisión, varía de universidad a universidad, por lo que el margen dentro del cual un aspirante tiene la posibilidad de ser aceptado al estudio, es bastante grande. Cada universidad tiene la libertad, además, de incluir otros criterios para la selección. De esta posibilidad se hace mucho uso, ya que la aplicación de tests de rendimiento escolar y de otros tests para diagnosticar habilidades no solamente garantizan una selección que será cabal, en el caso individual. Sin embargo, por medio de tests, la medición objetiva y comparable en una gran dimensión, forma siempre la base para la decisión.

BIBLIOGRAFIA

- BARTEL, H. 1961. Möglichkeiten und Grenzen einer objektiveren und damit gleichmäßigeren Notengebung und von besser vergleichbaren sowie als gerechter empfundenen Schulzensuren. *Die höhere Schule*, 14, 167-171 und 197-201.
- BERNART, E. 1962. Schulleistung und Begabung. München/Basel (= *Erziehung und Psychol.*, Nr. 19).
- BESSLER, E. 1927. Die Gleichförmigkeit der Noten in der Volksschule. *Arch. Ges. Psychol.*, 59, 1-74.
- BLUMENFELD, W. 1931. Urteil und Beurteilung. Leipzig (= *Arch. Ges. Psychol.*, 3. Erg.-Bd.).
- BOBERTAG, O. 1933. Leistungsschätzung und Leistungsmessung in der Schule. *Z. Pädag. Psychol.*, 34, 377-393.
- BOBERTAG, O. 1934. *Schülevaustese*. Berlin.
- CARTER, R. S. 1952. How invalid are marks assigned by teachers? *J. Educ. Psychol.*, 43, 218-228.
- CHAMPNEY, H. y MARSHALL, H. 1939. Optimal refinement of the rating scale. *J. Appl. Psychol.*, 23, 323-331.
- CONKLIN, E. S. 1925. The scale of values method for studies in genetic psychology. *Univ. Ore. Publ.*, 2, Nr. 1; cit. Guilford 1954.
- DIEDRICH, K. 1961. *Elternhaus und Schule*. Berlin/Neuwied.
- DÖRING, W. O. 1931. *Pädagogische Psychologie*. Osterwiesek.
- DOHSE, W. 1963. *Das Schulzeugnis*. Weinheim.
- EBEL, R. L. 1965. *Measuring Educational Achievement*. Englewood Cliffs.
- ENGELMAYER, O. 1960. *Psychologie für den schulischen Alltag*. München.
- FLITNER, A. 1966. Das Schulzeugnis im Lichte neuerer Untersuchungen. *Z. Pädag.*, 12, 511-538.
- FRANKE, W. 1922. Die Frage der Zensurierung. *Dt. Philologenbl.*, 30, 408-409.
- GAUPP, A. 1963. Überlegungen und Tatsachen zur Bewährung der Schülersause. In: Ingenkamp, K. H. *Pädagogisch-Psychologische Untersuchungen zum Übergang auf weiterführende Schulen*. Weinheim 55-80.
- GÖLLER, A. 1966. *Zensuren und Zeugnisse*, Stuttgart.
- GRASSAU, E. 1968. *Universitätsausleseverfahren in Chile*. In: Ingenkamp und Marsolek 1968, 765-787.
- GUILFORD, J. P. 1954. *Psychometric Methods*. New York.
- GUILFORD, J. P. 1965. *Fundamental statistics in psychology and education*. New York.
- HAKE, H. W. y GARNER, W. H. 1951. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. exp. Psychol.*, 42, 358-366.
- HARTOG, PH. y RHODES, E. C. 1936. *An examination of examinations*. London.
- HASEMANN, K. 1964. *Verhaltensbeobachtung und Verhaltensbeurteilung in der psychologischen Diagnostik*. Göttingen.
- HASEMANN, K. 1968. Begriff und Institution der Hochschulreife. Probleme der Ermittlung von Eignungsvoraussetzungen für die wissenschaftliche Berufsausbildung. *Habilitationsschrift*, Freiburg (Breisgau).
- HERMANN, E. 1948. Reform unseres Notensystems. *Pädagogische Welt*, 2, 737-744.
- HÖFLER, A. 1920. *Reifwerden, Reifmachen und Reife-Prüfen*. Leipzig.
- HOPP, A. D. y LIENERT, G. A. 1965. Eine Verteilungsanalyse von Gymnasialzensuren. *Schule und Psychologie*, 12, 139-150.
- HORNEY, W. 1960. Schülerbeobachtung und-beurteilung. In: Horney, W. et al. (Hgg.). *Handbuch für Lehrer*. Gütersloh. Bd. I, 150-166.
- HUTH, A. 1955. Tatsachen und Gesetze im Notewesen. *Berufspädag. Z.*, 4 45-47.
- INGENKAMP, K. y MARSOLEK, TH. 1968. *Möglichkeiten und Grenzen der Testanwendung in der Schule*. Weinheim.
- JUST, G. 1934. Zum Problem: Schulleistung und Lebensleistung. *Z. Ang. Psychol.*, 47, 161-172.
- JUST, G. 1939. Erbpsychologie der Schulbegabung. In: Just, G. (Hg). *Handbuch der Erbbiologie des Menschen*. Bd. 5, Tl. 1, 538-591.
- KAPUSTE, H. y SUND, M. 1966. Die Notwendigkeit der Verbesserung der Auslesekriterien. In: Arb. gemeinsch. Ausbild. forsch., Arb. gruppe Dr. Kapuste. Die Probleme der Zulassung zum Medizinstudium in der Bundesrepublik. *Deutsches Ärzteblatt - Ärztl. Mitt.*, 63, 1206-1213.
- KIESSLING, A. 1929. Leistungsbeurteilung und Leistungsbewertung. *Z. Pädag. Psychol.*, 30, 569-573.
- KLEINERT, H. 1951. Zensur, Zeugnis, Schulzeugnis. In: Kleinert, H., Stucki, H. et al. (Hgg.). *Lexikon der Pädagogik*. Bern. II. Bd., 919-923.

- KÖRNER, E. 1956. Leistungsbeurteilung nach Punkten. *Erziehung und Unterricht*, 106, 1-7.
- KOHAN, N. C. DE 1968. *Möglichkeiten und Grenzen für die Anwendung objektiver pädagogischer Tests in Argentinien*. In: Ingenkamp und Marsolek 1968, 141-145.
- LÄMMERMANN, H. 1927. Das Mannheimer kombinierte Verfahren der Begabtenauslese. *Beih. Z. Ang. Psychol.*, 40.
- LIENERT, G. A. y HOPP, A.-D. 1964. Über die Interkorrelationen von Gymnasialzensuren. *Schule und Psychologie*, 11, 193-206.
- LIETZMANN, W. 1931. Von der amerikanischen Testbewegung. *Z. Pädag. Psychol.*, 32, 473-486.
- LUCKER, E. 1965. *Die Berufswahlsituation eines Abiturientenjahrganges*. München/Basel.
- MARTINAK, E. 1900. Psychologische Untersuchungen über Prüfen und Klassifizieren. In: Mally, E. und Tumlriz, O. (Hgg.). 1929. E. Martinak, *Psychologische und Pädagogische Abhandlungen*. Graz. 157-174.
- MEESE, F. 1922. Teilung der Zensur "Genügend". *Dt. Philologenbl.*, 30, 500.
- MERZ, V. 1951. Das Problem einer rechten Schülerbeurteilung, Noten oder Wortzeugnis? *Die Pädagogische Provinz*, 5, 244-252.
- MOCK, A. 1964. Entwicklung, Korrelation und Faktorenanalyse der Zeugnisnoten von Oberschülern. *Dis. phil.*, Köln.
- ORLIK, P. 1961. Ein Beitrag zu den Problemen der Metrik und der diagnostischen Valenz schulischer Leistungsbeurteilungen. *Z. exp. ang. Psychol.*, 8, 400-408.
- ORLIK, P. 1967. *Kritische Untersuchungen zur Begabtenförderung*. Meisenheim/Glaun.
- PEDRAZA, F. G. DE, y BACA, A. 1965. Validación del equipo de pruebas para selección universitaria. *Rev. Psicol.* (Bogotá), 10, N^o 1, 75-78.
- PFÄFF, A. 1961. *Die Interkorrelation von Schulnoten an der Oberstufe neusprachlicher Gymnasien*. Marburg; cit. Mock 1964.
- PFANZAGL, J. ² 1964. *Allgemeine Methodenlehre der Statistik*. I. Berlin
- PIÉRON, H. 1963. *Examens et Docimologie*. Paris; cit. Göller 1966.
- REIRING, H. 1954. Noten- und Punktwertung. In: *Dt. Inst. f. wissenschaftl. Pädag.*, Münster und *Inst. f. Vergl. Erzieh. Wissenschaft*, Salzburg (Hgg.). *Lexikon der Pädagogik*, Freiburg i. Br. III. Bd. Sp. 692-694.
- Rektorenkonferenz, Westdeutsche. 1964. Empfehlungen der Arbeitsgruppe IV für die Plenarberatung über Notmaßnahmen angesichts der Überfüllung der Hochschulen. *Empfehlungen, Entschlüsse und Nachrichten* (Schwarze Hefte), Stück 77, 99-101.
- REUCHLIN, M. 1967. *Schul- und Berufsberatung*. Weinheim.
- RIEBESSELL, P. 1919. Die mathematischen Grundlagen des Zensurensystems. *Mitt. Mathem. Gessellsch.* Hamburg, 5, 269-273.
- RUCH, G. M. 1929. *The Objective New-type Examination*. Chicago; cit. Ebel 1965.
- RUPPERT, J. P. 1954. Die Schule als Sozialgebilde und Lebensform. *Sozialpsychologie im Raum der Schule*. Weinheim.
- SANDER, J. 1963. Schülerleistung und Lehrerurteil im Blickfeld der Statistik. In: Specht, K. G., Rasch, H. G. und Hofbauer, H. (Hgg.), *Studium sociale*. Köln/Opladen. 475-501.
- SCHOTT, E. 1920. Einheitszeugnisse und Zeugnisseinheiten. *Dt. Philologenbl.*, 28, 522-523.
- SCHOTT, E. 1922. Unsere Zeugnisbezeichnungen. *Dt. Philologenbl.* 30, 552-553.
- SCHREIBER, H. 1899. Gegen Prüfungen und Noten. *Z. Philos. Pädag.* 6, 31-38.
- SCHÜTZ, E. 1968. Studien- und Schulleistungen. *Deutsches Ärzteblatt*, 65, 510-516.
- SEELIG, G. F. 1967. Schullaufbahnenlenkung durch Leistungsmessung. "Educational Testing in America". *Der Schulpsychologe*, 14, H. 1, 21-24.
- SIMONEIT, M. 1952. *Fort mit der Schulzensur. Das Beurteilen von Schülerleistungen*. Berlin.
- SOST, J. 1926. *Wessen und Bedeutung der Schulzeugnisse und ihre pädagogische und psychologische Auswertung*. Paderborn.
- SPENGLER, G. 1967. Abiturleistungen und akademischer Examenserefolg. *Arbeit und Leistung*, 21, 157-167.
- STEVENS, S. S. 1951. Mathematics, measurement and psychophysics. In: Stevens, S. S. (Hg.). *Handbook of experimental psychology*. New York, 1-49.
- STROHMEYER, H. 1930. *Zeugnisse, Versetzungen, Prüfungen und Berechtigungsverfahren an den Höheren Lehranstalten*. Braunschweig.
- STRUNZ, K. ⁴ 1967. Die psychologische Diagnostik im Rahmen der Schularbeit. In: Strunz, K. (Hg.). *Pädagogische Psychologie für Höhere Schulen*. München/Basel. 124-143.
- SYMONDS, P. M. 1924. On the loss of reliability in ratings due to coarseness of the scale. *J. exp. Psychol.*, 7, 456-461.

- THORNDIKE, E. L. ²1912. *Mental and Social Measurements*. N. Y.
- TODT, E. 1966. Untersuchungen zur Vorhersage von Schulnoten. *Psychol. Forsch.*, 29, 32-51.
- TRAYERS, R. M. W. y GRONLUND, N. E. 1950. Meaning of Marks. *J. Higher Educ.*, 21, 369-374; cit. Ebel 1965.
- VICENCI, O. 1953. Die Bewertung der Schülerleistung. *Erziehung und Unterricht*, 103, 137-140.
- WACENSCHNEIN, M. 1954. Noten. *Die Sammlung*, 9, 411-414.
- WEIDIG, E.-R. 1961. *Die Bewertung von Schülerleistungen*. Weinheim.
- WEINGARDT, E. 1964. *Korrelation und Voraussagewert von Zeugnisnoten bei Gymnasialten*. München/Basel.
- WEINGARDT, E. 1967. Der Voraussagewert des Reifezeugnisses für wissenschaftliche Prüfungen. In: Merz, F. (Hg.). *Bericht über den 25. Kongr. DGfPsychol.*, Münster 1966. Göttingen. 679-684.
- WEISS, R. 1964. Die Berechnung einer "Schulleistungszahl". *Schule und Psychologie*, ii, 114-121.
- WEISS, R. 1965. *Zensur und Zeugnis*. Linz.
- WISSENSCHAFTSRAT (Hg.). 1967. *Empfehlungen zum Ausbau der wissenschaftlichen Hochschulen bis 1970*. Bonn.
- WÖLKER, H. 1968. *Zensuren aus dem Computer*. München.
- ZEIDLER, K. 1928. Zur Frage der Zeugnisgestaltung. *Die Erziehung*, 3, 175-183.
- ZILLIG, M. 1928. Einstellung und Aussage. *Z. Psychol.*, 106, 58-106.
- ZILLIG, M. 1959. Psychologische Probleme der Oberstufe der Volksschule. In: Hetzer, H. (Hg.). *Pädagogische Psychologie* (= Handbuch d. Psychol., Bd. 10). Göttingen. 273-292.