

# La importancia de la examinación en matemáticas: un enfoque sistémico

## The importance of mathematics examination: a systemic approach

*Omar Mejía Pérez\**

### RESUMEN

En la evaluación educativa existen diversos elementos que posibilitan, de mejor manera, la apreciación y el enjuiciamiento del desempeño escolar. En matemáticas, la naturaleza abstracta, deductiva y procedimental de las asignaturas demanda un proceso de verificación de los logros de aprendizaje tan riguroso como objetivo y justo. Uno de los recursos que demostró –mediante la intervención investigativa etnográfica y la aplicación de ejercicios experimentales– ser capaz de recoger información, interpretarla, analizarla, valorarla, enjuiciarla y calificarla de manera pormenorizada, objetiva y ecuanime fue la examinación sistémica. Esta debe valerse de otro instrumento de verificación para que procedimentalmente apoye al docente de manera eficiente a emitir juicios calificativos y tomar decisiones.

**Palabras clave:** matemáticas, examinación sistémica del aprendizaje, intervención investigativa etnográfica, educación media superior

### ABSTRACT

Within educational assessment there are various elements that enable better the appreciation and judgment of school performance. In math, the abstract, deductive and procedural nature of subjects requires a verification process of learning achievements as rigorous as objective and fair. One of the resources that showed –thorough ethnographic research intervention and the experimental exercises– capability to gather, interpret, analyze, value, judge and grade the information in an objective, equable and detailed way was the systemic examination. This examination must use another verification instrument in order to give teacher procedural support in an efficient way to qualify and make decisions.

**Keywords:** mathematics, systemic examination of learning, ethnographic research intervention, high school

## INTRODUCCIÓN

En los enfoques pedagógicos contemporáneos se ha estipulado con rigor que la educación debe ser de calidad; que la evaluación debe ser formativa, auténtica e integral, pero ¿qué implementar para lograrlo?, ¿cómo llevarla a cabo?, ¿cómo instrumentar un

\* Universidad Autónoma del Estado de Morelos y Centro de Estudios Superiores de México; klvita@hotmail.com

modelo de evaluación para valorar la matemática como un proceso de indagación, de apropiación de objetos, algoritmos y razonamientos de manera justa, objetiva e integral?

Las respuestas a estas preguntas no son sencillas. No obstante, uno de los pilares de la evaluación que es muy útil para estimar conocimientos factuales, apropiación de objetos y algoritmos matemáticos, procedimientos lógico-deductivos, problemas hipotéticos en un sistema educativo donde “la evaluación de los educandos comprenderá la *medición* en lo individual de los conocimientos, las habilidades, las destrezas y, en general, del logro de los propósitos establecidos en los planes y programas de estudio” (Cámara de Diputados del H. Congreso de la Unión, 2013: 18,<sup>1</sup> cursivas nuestras), es indudablemente el examen; pero no el corriente y tradicionalista, sino el examen como parte del arduo proceso evaluativo que pretende retratar, con basta aproximación, la realidad para valorarla y formular juicios útiles y objetivos. Este instrumento está inmerso en un sistema de múltiples factores y recursos que auténticamente pretende explorar, reconocer, estimar y validar esta compleja realidad educativa. Denominamos a este proceso concreto *examinación sistémica*.

Realizamos el estudio investigativo al interior de tres módulos de Educación Media Superior a Distancia (EMSAD) del estado de Morelos a cargo del Colegio de Bachilleres del Estado de Morelos (COBAEM). Fue producto de un proceso de indagación etnográfica basada en la propuesta investigativa de María Bertely (2000) y apoyada en tres ejercicios experimentales, que consistieron en la aplicación de exámenes a distintos grupos y su posterior análisis.

## LA NECESIDAD DE LA EXAMINACIÓN

¿Cuál es el verdadero sentido de evaluar en la escuela? Si bien entendemos que *medir* es solo una de las articulaciones institucionales necesarias para la selección, la estratificación y la promoción de los estudiantes, el complicado concepto de evaluar podríamos concebirlo como

<sup>1</sup> Artículo 50 de la Ley General de Educación, Última Reforma, *Diario Oficial de la Federación*, 10 de junio de 2013.



... un proceso de recogida de información, además de continuo, ordenado, sistemático y complejo que requiere de una dedicada planeación; que es en sí mismo un sistema de subprocesos alternos, de distinto orden pero articulados y que tienen como fin común la constatación y estimación de la construcción del aprendizaje y el desarrollo de facultades del alumno, en donde los resultados arrojados se interpretan con justicia y objetividad para tomar decisiones y actuar, en consecuencia, para incrementar dialécticamente el desarrollo integral de los educandos y su vida plena (Mejía, 2012: 35).

Entonces, desde esta perspectiva, el sentido legítimo de evaluar debería estar en función de tres aspectos: a) la recolección de datos fidedignos para su procesamiento, b) el análisis escrupuloso de estos, y c) la estimación de logros y elaboración de juicios de valor. Así, resulta tan verídico y evidente que para conseguir una evaluación *de calidad*—soportada en lo anterior como sus propios estándares— es necesario diseñar un sistema de evaluación que considere el acopio de información, su núcleo y su vértebra principal, la confección y la aplicación del instrumento.

La ciencia exacta es ciertamente caprichosa. La matemática, como ciencia abstracta, procedimental, lógica y compleja requiere una enseñanza perfectamente sistematizada, ordenada, objetiva, relacional y, a su vez, útil. Por tanto, su evaluación debe responder a estos mismos principios. De esta manera, el instrumento que puede ser capaz de recoger, evidenciar, validar y arrojar información fiel y útil para estimar objetiva y meritoriamente un nivel de avance académico en matemáticas es el instrumento examinador, pero no solo.

El examen ha sido en sí mismo, por excelencia, una herramienta de recogida de información, un formato de interpretación y verificación, un aparato calificador y un medio de jueceo y promoción escolar. Este hecho ha creado una actual antipatía hacia él, tanto de los sujetos evaluados como de los teóricos educativos, y con justa razón. El problema nuclear es que el abuso de su implementación tanto como su incipiente operatividad desbordaron su propia utilidad. Su diseño durante largos años ha tenido poca preocupación, tal que el docente lo ha elaborado con mucha



insuficiencia académica y ética, así como con incongruencia en lo recopilado y lo valorado.

Las nuevas perspectivas pedagógicas, que no le brindan gran crédito, lo deben procurar como un recurso útil en la matemática, precisamente para valorar en esencia los procedimientos algorítmicos, tanto abstractos como hipotéticos; la sustancia de la matemática.

Pero, ¿cuál es entonces la verdadera importancia de utilizar el examen? La matemática es un conjunto complejo de abstracciones y relaciones causales del comportamiento físico y geométrico del mundo, así como la base del raciocinio y la lógica formal; por ello, su aprehensión y verificación requieren procesos metódicos. La matemática es una actividad educativa, y la matematización escolar puede establecerse como horizontal o vertical (Treffer, 1978, en García, 2001), pero, *grosso modo*, debería ser parte de ambas. Por un lado, es necesario enseñar al alumno a transferir problemas reales a problemas matemáticos, y de igual forma a analizar problemas como modelos matemáticos abstractos que tienen su origen y epílogo en la razón pura. Resultaría sumamente calmoso que el alumno fuese descubriendo, a través de su mediación de la realidad, conceptos y procedimientos matemáticos ya establecidos que solo requieren su análisis y aprehensión para desarrollarlos en su propia mediación de la realidad.

De esta forma, el estudiante debe desarrollar competencias cognitivas que le permitan ejercitar y memorizar, adecuadamente, distintos algoritmos y modelos matemáticos que pueda traducir y utilizar en su realidad inmediata y mediata. La mejor forma de verificar el nivel de desarrollo competencial de esto es mediante un procedimiento capaz de arrojar información fidedigna sobre la manipulación de variables aritméticas, geométricas y algebraicas conforme al uso del análisis estructurado y la reflexión lógico-matemática: la examinación sistémica.

Efectivamente, el examen escrito es una prueba contundente que somete al examinado a un espacio y tiempo concretos y limitados, que ejerce una presión irrevocable y confina su contenido a ítems poco prácticos; sin embargo, también gracias a ello el examen puede ofrecer fidelidad subjetiva de introspección, verificación efectiva de la adquisición procedimental-memorística de



algoritmos y modelos matemáticos, y sentido de responsabilidad por el estudio.

Este instrumento, además, es el único que puede garantizar el cotejo fehaciente del nivel de desempeño mínimo requerido en determinado contenido a evaluar. El examen, a diferencia de otros instrumentos, puede arrojar información plenamente clara sobre la manipulación de procesos algorítmicos requeridos por los programas de estudio, y ayudar con mucha facilidad a estimar la reprobación o la acreditación de un sujeto en un determinado momento didáctico, con base en los criterios de desempeño mínimos de un curso.

Como uno de los ejercicios experimentales que utilizamos dentro del proceso etnográfico al interior de los módulos de EMSAD, se aplicó un examen escrito al grupo de tercer semestre donde se requería lo siguiente:

Un trapecio como se muestra tiene de  $B = 12$  cm, de  $b = 10$  cm y de  $h = 4$  cm. Si la fórmula para determinar el área del trapecio es  $(B + b) \times h / 2$  y el perímetro de cualquier figura se calcula sumando todos sus lados, determina el área del trapecio y su perímetro. De no poder hacer uno de los cálculos por falta de datos, escribe qué falta. Checa que la diagonal del trapecio es una HIPOTENUSA que al acercar la altura ( $h$ ) al vértice de arriba de cualquier lado se forma ¿Ya la viste? Ahora solo debes usar el teorema de Pitágoras para sacar el lado y luego poder obtener el perímetro.



El examen se aplicó a cuatro pequeños grupos de siete alumnos para evitar distracciones y plagio; cada grupo tuvo media hora para resolverlo. Siete lo resolvieron correctamente –algunos con cierta ayuda del asesor–; seis se acercaron a la respuesta relativamente, cinco implementaron diversos algoritmos por los que era casi imposible llegar a las respuestas correctas y diez demostraron no poder implementar algún algoritmo relacionado con el problema.

Se eligieron al azar dos alumnos de cada subgrupo resultante –de acuerdo con los resultados que acabamos de señalar–. Se les entrevistó conforme a la siguiente guía:

1. ¿Entiendes qué es el teorema de Pitágoras?, ¿cómo se enuncia y para qué sirve?
2. El examen que aplicaste, ¿para qué te sirvió?
3. ¿Crees que en tu vida futura pueda serte útil?

Los dos alumnos del primer grupo –que resolvieron correctamente el examen– entendieron lo que es el teorema, cómo se enuncia y para qué sirve; entendieron también que el teorema era fundamental para resolver el problema y consideraron de manera importante que es de suma utilidad. Uno de los del segundo grupo supo qué era el teorema, cómo se enunciaba y para qué servía, el otro solo supo cómo se enunciaba. El primero explicó abundantemente cómo lo había aplicado en su examen; el segundo indicó que lo había aplicado pero que no lo entendía bien –lo aplicó solo en una parte de la figura–. Ambos dijeron que les sería muy útil en su vida futura. Del resto de los alumnos solo uno del tercer grupo supo qué era el teorema y para qué servía. Nadie lo aplicó en el examen; el que sí lo supo manifestó: “no manche’, ¿a poco así se resolvía?”. Todos dijeron que sí les sería útil en la vida futura, aunque uno del cuarto grupo no fuera convincente.

Esto sugirió que los alumnos que son capaces de resolver un examen escrito de matemáticas adquieren el desarrollo suficiente de los desempeños, como en este caso la distinción, el manejo y la aplicabilidad del teorema de Pitágoras. Los que no lo pudieron resolver, evidentemente no lo adquirieron. Ahora, debiera ser notorio que el instrumento examinador nos permitió establecer con certeza que fue capaz de recoger información suficiente, fiel y contundente sobre los alumnos en torno a la adquisición del algoritmo pitagórico y su racionalización para trasponer a contextos geométricos no triangulares, gracias a que, luego de contrastarlo con la entrevista, logramos verificar quién lo había aprendido bien, quién en un grado menor y quién definitivamente no, lo que coincidió, en concreto, con los resultados arrojados en el examen.

Díaz Barriga (1993) comenta que Horace Mann distingue varias ventajas en los exámenes escritos: a) se pueden lograr más elementos de juicio sobre el rendimiento de cada alumno, b) se dispone de una constancia escrita de ese rendimiento, c) a cada



alumno se le formulan las mismas preguntas; de ese modo todos reciben igual trato, y d) hay menos posibilidades de favoritismo o de animadversión respecto de un alumno o maestro concreto.

Barajas y Gilio (2009) en el X Congreso Nacional de Investigación Educativa celebrado el 25 de septiembre de 2009 en Veracruz, presentan varios resultados relevantes de su investigación, “El examen dentro de un proceso de evaluación participativa: una experiencia en matemáticas de bachillerato”. Entre sus hallazgos más significativos destacan:

- Los estudiantes manifiestan sentimientos de rechazo hacia el examen en general, en particular frente al de matemáticas ya que es muy difícil; paradójicamente, lo consideran un instrumento útil para evidenciar sus aprendizajes.
- La comunicación clara y la negociación respecto a la evaluación entre profesor y estudiantes generan una disminución de ansiedad en algunos alumnos frente al examen escrito.
- El examen en cuanto práctica educativa y, por ende, práctica social es una actividad que conlleva significaciones para los diferentes actores involucrados y que “cobra sentido en el marco de una configuración socio-cultural determinada” (Vain, 2003, en Barajas y Gilio, 2009).
- Se ha conformado toda una pedagogía del examen. Se pretende que a través de la mejora técnica de los exámenes también mejore la educación en su conjunto.

Efectivamente, la examinación escrita en la actualidad presenta múltiples deficiencias, variantes y tensiones nerviosas; no obstante, es la herramienta apropiada que puede proporcionar fidelidad y transparencia en la recuperación de los procesos matemáticos, siempre y cuando se planea y ejecute con sistematización, ética y conciencia matemática. Por supuesto, hay un resto de instrumentos que son un complemento muy importante en la gran mayoría de los contenidos de los programas, pero la examinación sistémica es el recurso más acertado para verificar el desarrollo real individual de competencias matemáticas abstractas.



## CONFECCIÓN DE EXÁMENES ESCRITOS

La cautela, la dedicación y la intención para diseñar un examen son su principal sustento de eficacia. Los exámenes, como ya hemos mencionado, han sido objeto de menoscabo debido a la poca atención en su confección. Uno eficiente jamás podrá ser un instrumento elaborado casualmente, sin objetivos claros o de manera desdénosa, sino con mucha objetividad, pertinencia y claridad.

Desde el conductismo, un examen es “cualquier medio que se usa para medir el rendimiento del alumno” (Rodríguez y García, 1990). En la pedagogía contemporánea, un examen debería ser *un instrumento que permita verificar el progreso del desempeño individual de un sujeto a la luz de determinadas competencias*. Así, un examen debe ser un utensilio afinadamente planeado, con objetivos congruentes a los contenidos revisados, una selección escrupulosa de reactivos y un plan de *verificación de información* basado en el diseño de un instrumento que viabilice el análisis de los datos, evitando que el examen, de manera automática, arroje valores rudimentarios e imprecisos en su revisión.

De este modo, un examen, para fungir como un instrumento verdaderamente útil, debe ser sometido a un proceso riguroso de interpretación, conforme a otro instrumento que facilite el ordenamiento de los datos y permita la clarificación del avance de acuerdo con los criterios previamente sentados; es decir, un *instrumento de verificación*.

Un docente que revisa el desempeño *reconoce los diferentes tipos de ecuaciones de la circunferencia y las transforma de una forma a otra*, y necesita examinar el progreso del desempeño conforme al análisis de problemas en el plano de una circunferencia con centro fuera del origen para trabajar la ecuación canónica; conviene diseñar un reactivo en el que se enuncie: “obtén la ecuación canónica de la circunferencia que tiene su centro en el punto (4,-3) y pasa por el punto (0,-1). Dibújala en el sistema de coordenadas y resalta el radio que se forma con el punto anterior”.

Como es posible observar, el reactivo contiene una fuerte carga analítica para la primera requisición<sup>2</sup> (obtén la ecuación de la

<sup>2</sup> Nos referimos con *requisición* en el área didáctico-evaluativa a la acción y efecto de solicitar la atención o resolución de un reactivo en un instrumento examinador.



circunferencia), y cargas de menor proporción para las dos restantes (dibújala en el sistema de coordenadas y resalta el radio). El docente debe valorar las tres requisiciones en diferente proporción, pues la primera involucra mayor complejidad cognitiva y racional –saber cómo interpretar la información instruida, cómo enunciar la ecuación y qué elementos están involucrados, y cómo manipular los datos para obtener los valores de las incógnitas–, además de ser el principal factor que constituye el desempeño requerido, mientras las otras dos solo necesitan conocimientos previos básicos sobre el plano cartesiano y geometría analítica, y las bases generales de la circunferencia.

Al desarrollar el ejercicio se puede ver que para resolverlo es necesario utilizar el teorema de Pitágoras en el plano para determinar la longitud del radio y con ello la ecuación. De esta forma, el instrumento de verificación, en este caso una hoja de escalas estimativas, podría valorar lo siguiente:

- a) Interpretación, análisis y planteamiento del problema (30%).
- b) Implementación del teorema de Pitágoras (30%).
- c) Construcción de la ecuación (20%).
- d) Esquemmatización de la circunferencia y el radio (20%).



El instrumento de verificación también puede arrojar solo criterios verbales o puntuaciones –en el primer caso, enunciados como: “pleno, suficiente, insuficiente”, o mejor aún, más completos como: “implementa el teorema sin errores”, o “implementa el teorema con algunos errores”; y en el segundo caso, “si el primer rubro de la escala se cumple con amplia satisfacción, acumula cinco puntos o decrece cero puntos”; si el segundo se cumple con algunas observaciones, “incrementa cuatro puntos o resta uno”, y así sucesivamente–. La cuestión más importante de todo esto es poder notar que los exámenes son excelentes instrumentos que, aunados a otro, pueden sistematizar en gran medida, sin ser decisiones frívolas e inflexibles, el proceso complejo de recoger información, interpretarla y valorarla en aras de fortalecer la ecuanimidad educativa.

Ahora bien, los exámenes tienen dos partes que es necesario atisbar con cautela, partes que lo determinan esencialmente y que logramos recuperar dentro del espacio etnográfico:

- a) Su estructura de contenido.
- b) Los ítems o reactivos que lo conforman.

Estas dos piezas son las que pueden llegar a avalar la pertinencia, la valía y la confiabilidad del instrumento de manera genuina. Aunque la interpretación de la información colectada y su traducción escalar son procesos que también afianzan la ecuanimidad evaluativa, los primeros, al ser parte intrínseca del instrumento, son aquellos en los que se deposita primariamente el crédito fidedigno de la captura de los datos de los sujetos examinados.

La estructura de un examen es la organización que se planea en él. Básicamente la componen las instrucciones generales, el orden y el número de apartados, el tipo de reactivos y la distribución espacial de su contenido; esto es, su arquitectura medular.

Aunque parezca soso, la estructura de un examen define de buena forma la eficiencia del instrumento. Las instrucciones deben ser muy claras y precisas, esto es, deben ser plenamente decodificables por los sujetos evaluados; una confusión mínima puede desembocar en una resolución incorrecta de todo el examen. La etnografía también nos permitió encontrar que la mayoría de los alumnos ha desarrollado, improductivamente, el hábito de no leer las instrucciones. Solicitamos a un docente que escribiera, al final de las instrucciones generales de un examen que aplicó al quinto semestre, la leyenda “no resuelvas este examen, entrégalo sin contestar ninguna pregunta y obtendrás 10, siempre y cuando no le digas a nadie y salgas del salón inmediatamente”. Solo tres alumnos de 28 lo hicieron. Esto nos invita a pensar en que hay que resaltar las instrucciones de alguna manera, exhortarlos a leerlas y forjarles el hábito de hacerlo. Algo tan simple puede mejorar el aprovechamiento académico.

El número de apartados y su ordenamiento pueden facilitar o complicar la resolución de un examen, más aún cuando este tiene una secuencia de resolución. El número de apartados debe dise-



ñarse concienzudamente. En el tradicionalismo, con frecuencia, se diseñaban exámenes con varios apartados de distintos tipos de ítems; ejemplo: I) relaciona columnas, II) selecciona la respuesta correcta, III) escribe F o V, IV) completa el espacio en blanco de la oración, V) responde las preguntas abiertas. Y aunque este modelo sigue siendo utilizado, los exámenes generalmente tienden a ser menos diversificados en sus apartados.

En matemáticas es preciso definir los desempeños y las competencias que se valorarán. Esto no significa que en otras materias no, sino que, en particular, en matemáticas, que se pretende profundizar en el nivel de análisis y razonamiento deductivo, es mayor el énfasis necesario en la estimación de requisiciones. Si se requiere valorar la adquisición y el manejo de un algoritmo determinado en algún tema, será difícil hacerlo si el apartado solicita la relación de columnas, ya que la abstracción metodológico-verbal dista mucho de su aplicación.<sup>3</sup> En cambio, si lo hago planteando un problema donde sea necesaria su implementación, se podrá valorar con muchos más elementos el manejo del algoritmo, que ha adquirido y que le falta. Esto nos lleva, inmediatamente, a la selección y el diseño de los ítems.

Un ítem o reactivo es un eje temático que inquiera en sí mismo resolución; “es aquello que produce una respuesta o reacción, como efecto de un estímulo” (Castañeda, 2008). Puede ser una pregunta, un eje interrogativo, una requisición definida o un enunciado incompleto. Los apartados de un examen contienen ítems que definen el tipo de estos. Los ítems tienen, generalmente, una carga semántica muy importante que define la facilitación o el impedimento de su resolución.

<sup>3</sup>Cuando aplicamos el instrumento arriba descrito a otro grupo como un segundo ejercicio experimental, le adicionamos otro reactivo igual, pero con otros datos y cuatro respuestas de selección múltiple; pudimos notar una amplia incongruencia en la apreciación de lo que los alumnos habían demostrado saber. El 16% resolvió ambos ejercicios procedimentalmente correctamente, pero el 40% el segundo, de selección múltiple. Esto indica que el 24% oscilante no es capaz de resolver un problema procedimentalmente, pero sí de seleccionar una respuesta correcta a un problema. Como podremos observar, la diferencia sería inmensa si tomáramos el reactivo 1 o el 2 para valorar y calificar. Cuando entrevistamos a los alumnos del 24% y solicitamos que resolvieran un nuevo ejercicio similar, notamos que todos presentaron dificultades para la resolución de aquellos de aritmética y geometría elemental, tanto como de razonamiento, deducción y procedimiento. Demostraron no ser suficientemente capaces de hacerlo, no como indicó el reactivo 2.



Muchas veces los diseñadores de exámenes asumen que el código lingüístico implementado en sus ítems es plenamente decodificable por sus examinados. “El vocabulario utilizado para enunciar la pregunta es un aspecto vital. No importa si esta es escrita u oral... se debe conocer el significado del término o los términos de las palabras clave que especifican lo que requiere la pregunta” (Medina-Díaz y Verdejo-Carrión, 2001). “Si un Jet F-311 vuela a 2000 km/h de suroeste a nororiente y tarda media hora en arribar al hangar, ¿cuál es su velocidad promedio?”. El punto es claro, el lenguaje algunas veces no es conocido por los alumnos, y varios de ellos preguntan lo que no entienden, pero no todos. Si desde un principio, en el momento de la examinación, instruyo a mis alumnos para que todas las dudas me las comenten en mi escritorio, los que son introvertidos y que generalmente no preguntan nada preferirán quedarse con las dudas lingüísticas y así tratar de resolver el problema, aunque esto les signifique no tener clara la idea de resolución. Además, si no existe ninguna justificación fehaciente del porqué implementar un posible código desconocido en un ítem, de igual manera es absurdo e improductivo.

Como diseñadores de exámenes debemos procurar construir ítems que sean plenamente inteligibles; que las instrucciones y las requisiciones faciliten el desarrollo de su resolución y no que los complique. López (2008) lo llamaría “calibrar los reactivos”.

Como lo mencionamos arriba, la congruencia de la evaluación debe plasmarse ampliamente en los exámenes. Desde el enfoque por competencias es muy disfuncional tratar de valorar el desarrollo de una competencia simplemente por sus características de memorización. El vasto contenido algorítmico del área matemática implica la adquisición lógico-procedimental de muchos procesos, y su valoración restringe de cierto modo sus ítems.

La forma más viable de evaluar la mayoría de los contenidos de los programas, en cuanto a producto y procedimiento se refiere, es mediante ejercicios abstractos o problemas hipotéticos. Ello permite la abstracción evaluativa del examinador para poder analizar los procedimientos implementados y tener mayor amplitud de su valoración. No estamos en desacuerdo con utilizar otros tipos de ítems –de selección múltiple, de relación, etc.–, sino

que, estrictamente, estimamos más conveniente definir con mucha puntualidad los objetos de evaluación en un examen –que después serán cotejados con el instrumento de verificación–, y que ellos mismos determinen el tipo de ítem a utilizar, donde recomendamos mucho el uso de ejercicios abstractos y problemas hipotéticos. Hay un trecho abismal entre solicitar el desarrollo y la resolución de un problema y solo pedir la selección de una respuesta de entre tres o cuatro posibles que, además, muchas veces están diseñadas con una extrema obviedad para descartar opciones o con demasiada amplitud interpretativa.

Pitágoras era de origen:

- a) Mexicano      b) Griego      c) Báltico      d) Venusino

O bien:

El teorema de Pitágoras enuncia que:

- a) La suma de los cuadrados de los catetos es igual al cuadrado de la hipotenusa.
- b) La suma de los cuadrados de los catetos en un triángulo rectángulo es igual al cuadrado de su tercer lado.
- c) La suma de los cuadrados de los catetos de un triángulo es igual al cuadrado de la hipotenusa.
- d) La suma de los cuadrados de los catetos en un triángulo rectángulo es igual al cuadrado de su hipotenusa.



En ambos casos, como ocurre con frecuencia, los ítems están tan disfuncionalmente diseñados que convierten el examen en un instrumento de recolección de datos irrelevantes y de poca fidelidad. En el primer caso, el grado de complejidad es tan bajo que rebasa la permisibilidad de abstracción y competencia. En el segundo, la complejidad se eleva tanto que se vuelve un problema de interpretación y no de carácter matemático. Este tipo de ítems no solo propician la elección azarosa de la respuesta, sino que obligan a la interpretación inmediata de la valoración y generan un instrumento con poca veracidad y una limitada gama de opciones para calificar.

Como habrá de entenderse, el área de matemáticas requiere exámenes mucho más prácticas para la mayoría de sus contenidos. La adquisición memorística de sus elementos es poco útil y tiende a ser efímera. Resulta muy pertinente confeccionar los exámenes con apartados que no necesiten demasiado tiempo de atención, ítems claros y congruentes, y objetivos minuciosamente planeados para provocar exámenes productivos.

### **INTERPRETACIÓN DE LA INFORMACIÓN, EMISIÓN DE JUICIOS Y PONDERACIONES**

El hecho de asegurar la recolecta fidedigna de la información no garantiza la satisfacción del proceso examinatorio. Como docente examinador es preciso saber manipular la información recopilada, entenderla y procesarla para obtener de ella datos relevantes que permitan establecer juicios valiosos. “Reconstruir la realidad desde una perspectiva crítica no solo supone su contemplación aséptica y su descripción rigurosa. Es necesario interpretarla, saber dónde están las raíces de los comportamientos, las causas de las actuaciones, los efectos del desarrollo curricular (tanto explícito como oculto)” (Santos, 2008).

La interpretación de la información parte de la recogida de datos. Interpretar la información recogida significa decodificarla; traducir los datos recopilados para estar en una postura más objetiva y abundante de la situación competencial lograda. Esto implica conjuntar los hechos aislados que tuvieron lugar en distintos momentos y formas evaluativas. Uno no interpreta un producto sin tomar en cuenta las situaciones que le antecedieron. Definitivamente, la interpretación es la etapa más compleja de evaluar. Como establece su definición la Real Academia Española (RAE), interpretar es “explicar el sentido de alguna cosa según la opinión personal”. Esta subjetividad limita la interpretación como una actividad individual humana; ello conlleva a complementar este proceso con evaluaciones de los mismos sujetos evaluados u otros actores, que más tarde abordaremos.

De cualquier modo, interpretar es una acción que el docente evaluador realiza cotidianamente: interpreta códigos escolares, comportamientos, signos verbales, actitudes y, por supuesto,



datos diversos del entorno didáctico y de cada sujeto para ser evaluado. La actividad de la interpretación es una tarea docente implícita en su quehacer ético; junto con la de calificar, es una de las cargas responsivas más pesadas. Su escrupulosidad y su sistematización le permitirán tener más posibilidades de objetivar las situaciones competenciales dadas.

Los instrumentos de verificación permiten el acceso a una interpretación más panorámica y justa. Estos instrumentos se han venido utilizando formalmente desde la concreción del constructivismo curricular, pero han tomado mucho revuelo en la actualidad pedagógica. Resulta prudente insistir en la importancia que radica en ello. Como explicábamos, el proceso de interpretación es un acto subjetivo que denota la codificación de los datos recogidos; su regulación sistemática y su meticulosidad lo convertirán en un proceso de mayor veracidad y atenuarán la simpleza mecánica de la calificación inmediata de los exámenes ordinarios y la imprecisión de las impresiones superfluas de los productos escritos.

Los instrumentos de verificación han logrado, en la actualidad, importantes avances en la interpretación y la valoración de productos escritos. Tanto las matrices de rúbrica, como las listas de cotejo y las hojas de escalas estimativas han registrado resultados importantes para acentuar los aspectos evaluados de productos que anteriormente solo eran valorados desde lo empírico. Sin embargo, el examen necesita un procedimiento similar.

En la búsqueda de un método que fuera factible para interpretar la información de los exámenes escritos del área matemática, encontramos que los instrumentos de verificación son herramientas sumamente útiles para ello. Arriba establecíamos que los instrumentos de evaluación tienen dos naturalezas: de recolección y de verificación. Esta fuerte divergencia estipula no solo su naturaleza prominente sino, además, su función específica. Hemos intentado corregir este grave error operativo de la evaluación y demostrar que es tan útil como necesario.

Todo instrumento de recolección de información requiere un trabajo concienzudo para su lectura. Ello aplica desde el principio de los instrumentos apropiados para la evaluación, que contengan objetivos claros y acordes con el contenido curricular revisado. El examen, que encaja perfectamente en este rubro, no puede



ser de ningún modo un instrumento multifuncional. Desde el tradicionalismo ortodoxo, un examen, por su estructura, era capaz de interpretar y calificar casi al mismo tiempo. Sin embargo, su confiabilidad, incluso conforme a los preceptos tradicionalistas, era muy pobre.

Un examen podía solicitarle al alumno que seleccionara la respuesta correcta de: “el cuadrado del binomio  $(x-3)$  es:

- a)  $x^2+6x+9$     b)  $x^2+6x-9$     c)  $x^2-6x+9$     d)  $x^2-6x-9$ ”.

Además de que el ítem<sup>4</sup> es criticable por su estructura semántica y la claridad de su requisición, resulta imposible valorar el proceso de resolución del ejercicio y además posibilita, enormemente, llegar a una resolución azarosa. ¿Qué nivel de confiabilidad proporciona una prueba así?, ¿es factible?, ¿qué pretende valorar en sí? Si lo que pretende valorar es la adquisición del algoritmo de la potencia de un polinomio, el cuadrado de un binomio o la obtención de un trinomio cuadrado perfecto, el nivel de confiabilidad conseguido para valorar cualquiera de esos aspectos con ese reactivo es sumamente bajo; tiene una probabilidad de acertar de 25% mediante un método de selección casual; como lo único que varía de las respuestas es el signo, el problema concentra su carga resolutive en la manipulación de signos, por lo que jamás pretende valorar potencia de números o literales, multiplicación entre términos o ni siquiera simplificación del tetranomio por semejanza de términos, sino la operatividad entre signos. El examinado puede obtener un resultado completamente distinto a todos los proporcionados, pero lo único que deberá contrastar es el acierto de los signos.

Este examen, como los muchos de los que se aplican en la actualidad, arrojan inmediatamente una calificación escalar. De todos los asesores de matemáticas en EMSAD del estado de Morelos, ninguno aplica otro instrumento para verificar su examen –salvo nuestro caso–. De preferencia, el diseñador busca corresponder el número de ítems en múltiplos o submúltiplos de diez, ajustados

<sup>4</sup> Extraído de un examen recuperado del CEBTis 232 de Tetecala, Morelos, con fecha de aplicación del 15 de noviembre de 1994.



a la escala oficial; de ser preferible, se busca que el número de reactivos sea de diez para que cada ítem respondido de manera correcta equivalga a un *acierto*, lo que corresponde a un punto en la escala. Así, un examen con siete *aciertos* de diez equivaldrá a un siete de calificación. Este número que sencillamente es un indicador de aciertos, no podría convertirse, por ningún motivo, en una calificación que representase la valía de un desempeño académico. No al menos de uno que pretenda denotar justicia, mérito y objetividad de un proceso complejo de todo un momento evaluativo y educativo.

El meollo de este apartado no estriba en la deficiente elección y construcción del ítem —eso ya lo discutimos arriba—, sino en la insuficiente información que el ítem permite recoger. Por supuesto, ello limita mucho la interpretación.

Los exámenes deben ser recursos que permitan al alumno demostrar sus desempeños y competencias con los menores límites posibles. Eso provocará que la información obtenida mediante este recurso sea abundante y fecunda. De este modo, el evaluador tendrá la oportunidad de cotejar con mayor prudencia y minuciosidad los datos obtenidos y los logros estimados.

Para que esto pueda llevarse a cabo satisfactoriamente es muy importante planear los instrumentos —de recogida y verificación— de modo simultáneo. El diseñador debe establecer primero sus objetivos de manera muy puntual para elaborar el contenido del instrumento examinador y, a su vez, el instrumento verificador, de tal forma que ambos obedezcan a los objetivos convenidos y uno complementemente al otro.

Revisemos un examen que aplicamos, el cual enunciaba: “obtén la ecuación canónica de la circunferencia que tiene su centro en el punto  $(4,-3)$  y pasa por el punto  $(0,-1)$ . Dibújala en el sistema de coordenadas y resalta el radio que se forma con el punto anterior”.

Este ítem contiene tres requisiciones: la ecuación de la circunferencia, la circunferencia trazada en el plano y el radio formado. El proceso de interpretación cobra mayor sentido en un recurso planeado y diseñado con objetivos, que de una manera informal y poco sistemática. Inicialmente, el intérprete debe centrar su atención en el primer rubro; debe verificar que el examinado consiguió la interpretación, el análisis y el planteamiento del pro-





blema de manera satisfactoria, o desde nuestro ámbito disciplinar, matemáticamente correcta. De no ocurrir esto al pie de la letra, debe verificar el grado de consecución de manera cualitativa (con juicios verbales que caractericen el desempeño; esto es muy importante). Es necesario que analice los procedimientos acertados (interpretación, análisis y planteamiento) y registre cuáles sí consiguió satisfactoriamente y dónde estriba su error o errores. De ser necesario (es una recomendación que, entendemos, implica mayor esfuerzo, pero se agudiza y se agiliza con su práctica), el docente necesita descomponer los procesos en subprocesos menores e inferir la génesis de su error cuando no se evidencie de manera inmediata (por ejemplo, si demuestra interpretación, análisis, pero el planteamiento contiene inconsistencias, se puede verificar si responden a un error de lectura, de transcripción, de síntesis o de metodología). De ahí el intérprete podrá estimar *a priori* juicios de lo que sería la primera fase de interpretación, e incluso podrá ya elaborar sus primeras tasaciones porcentuales. Después debe continuar haciendo la revisión de los demás rubros basado en estos principios y, por último, será capaz de fabricar una estimación holística de todo el proceso desarrollado y el logro conseguido del sujeto examinado. Si fue emitiendo valores porcentuales en un primer momento, es posible sumarlos y cotejar el resultado con la apreciación global del desempeño para verificar si la apreciación final coincide con la sumatoria estimada parcial que valoró en cada rubro. De ahí solo tiene que empatar juicios estimativos verbales con una calificación, preferentemente porcentual.

## CONSTRUCCIÓN DEL INSTRUMENTO DE VERIFICACIÓN

La matriz de rúbrica o simplemente rúbrica es uno de los mejores instrumentos para *verificar* –más allá de solo cotejar– los datos recopilados en un instrumento de acopio. “La estrategia más recomendada para juzgar el mérito de una pregunta [propia-mente de un ítem o un instrumento de recogida de información en nuestro caso; acotación nuestra] para elaborar la respuesta, consiste en la utilización de una rúbrica para su revisión y ca-

lificación” (Medina-Díaz y Verdejo-Carrión, 2001). Por ser el instrumento de verificación más completo amplía las posibilidades de dimensión, construcción y especificación de categorías o estándares, rubros e indicadores. “La rúbrica constituye una guía para establecer qué aspectos van a ser considerados en la revisión y calificación” (*idem*).

Toda construcción de instrumentos evaluativos implica un esfuerzo metódico. El diseño de la matriz de rubros requiere mucha atención en sus unidades primordiales. Cuando debemos diseñar una rúbrica para verificar un examen se vuelve necesaria una metodología que salvaguarde la integridad y la legitimidad de todo el proceso de lectura, interpretación y validación de resultados, así como de la misma determinación de la calificación. Esta metodología debe sujetarse a los procesos lógicos de observación, escrutinio, reconocimiento y asimilación de datos que serán sometidos a una valoración. El acostumbrado proceso de “palomeo” y “tacheo” de ítems y su automática designación calificativa (numérica) deben sustituirse por un proceso lógico de lectura interpretativa, identificación de demandas y niveles de logro, valoración de progresos, estandarización de escalas, clasificación de resultados y determinación de desarrollos competenciales. De este modo, es posible pensar en un mecanismo que optimice el propio proceso de evaluación definiendo con claridad cada una de sus partes y sistematizando su metodología.

La construcción de la rúbrica, entonces, no es un procedimiento simple. Debe ser un instrumento que contenga a detalle todo lo relacionado con la interpretación, la valoración y la calificación de un examen. Nuestra experiencia etnográfica nos permitió encontrar que la mejor manera de elaborar la rúbrica era diseñándola simultáneamente con el examen. Cuando se confecciona el examen se planean las requisiciones y se visualizan con claridad los objetivos; ese es un excelente momento para continuar con el diseño de la rúbrica donde deberán contemplarse, congruentemente, tanto objetivos como requisiciones reales y replantear, de ser necesario, ajustes en ambos instrumentos, con el fin de lograr una asidua compatibilidad entre el acopio palmario, la valoración objetiva y la ponderación calificativa meritatoria.



Para construir el instrumento verificador resulta necesaria la detección de sus etapas; su empleo riguroso garantiza un instrumento confiable en todas sus dimensiones. Las etapas que sugerimos para la confección de una rúbrica, como producto de nuestro quehacer etnográfico bajo la confección y la aplicación de los ejercicios experimentales que desarrollamos en los módulos, son las siguientes.

### **Interpretación de los ítems e identificación de requisiciones**

La estructura semántica de los ítems guarda una connotación importante: define en esencia lo que estos solicitan, el “qué”, “cómo” y “en qué condiciones”. Cuando establecíamos la importancia de diseñar un instrumento examinador con reactivos precisos, nos referíamos justamente a esta parte. Sin embargo, no solo el alumno debe decodificar con puntualidad la requisición de los ítems, sino también el evaluador, pues es quien debe, como primer paso, detectar y comprender la naturaleza del reactivo: su valor, grado de complejidad, el esfuerzo que se requiere para su atención y el significado que representa dentro de los objetivos de desarrollo competenciales. A menudo podría pensarse que este paso puede omitirse en el entendido de que, al diseñar el examen, estos elementos son plenamente contemplados. No obstante, no ocurre así. A lo largo de los ejercicios experimentales encontramos que debíamos volver a valorar la naturaleza de cada reactivo para mantener fresco su propósito, dimensión y valor competencial.

### **Establecimiento de rubros**

Una vez que se han detectado las requisiciones efectivas, se debe construir los rubros o criterios de la matriz, que devienen de las requisiciones de los ítems. Su diseño debe estructurar semánticamente el objeto nuclear de los resultados esperados, es decir, el fragmento competencial, desempeño u objeto de aprendizaje específico de un logro deseado. Por ejemplo, concepto de función, en el reactivo: define lo que es una función matemática.



En nuestra práctica evaluativa reciente, los rubros se construyen en dos etapas: la primera es la detección inmediata de las requisiciones por ítem, es decir, clasificamos y construimos los rubros primarios, uno por ítem; esto agiliza la estructura de la rúbrica una vez que han sido identificadas con plenitud las requisiciones en cada reactivo. La segunda es la segmentación de los rubros primarios en rubros específicos, que son aquellos que precisan un fragmento competencial, desempeño u objeto de aprendizaje concreto, lo que significa que los rubros primarios se fragmentan tanto como sea necesario para individualizar las requisiciones. Por ejemplo, si el reactivo fuera: “define lo que es una función y una relación matemáticas”, el rubro primario podría ser “concepto de función y relación”, que tendría que segmentarse en dos: “concepto de función y concepto de relación”. Aunque el ítem los relaciona por su naturaleza de análisis matemático –pues esa es la intención de tenerlos en un solo reactivo–, se debe valorar por separado, debido a que el alumno puede manejar muy bien un concepto pero no el otro y, definitivamente, por la estructura curricular de la asignatura de matemáticas IV –que es donde se revisa este contenido–; es más importante conocer y manipular el concepto de función que el de relación, por lo que deberá considerarse de mayor peso el concepto de función cuando se establezcan los puntajes ponderativos.



### **Dimensión ponderativa porcentual**

Cuando se han establecido los rubros es muy importante dimensionar el peso porcentual que mantendrá cada uno a la hora de estimar los valores nominales y escalares. Esto significa que las estructuras de los exámenes sostienen porcentajes ponderativos distintos en cada segmento de ellos, es decir, cada reactivo tiene un valor referencial entre sí, que está en función del fragmento competencial, desempeño u objeto de aprendizaje, su grado de complejidad y la dimensión del esfuerzo requerido para su atención. Desde luego no podría dividirse una ponderación porcentual en cuatro partes iguales (25% cada una) en un examen que tuviera cuatro ítems así:

1. Define lo que es una función.
2. Define lo que es una relación.
3. Explica qué es una función inversa y elabora un ejemplo.
4. Tabula, grafica y clasifica en todas las categorías revisadas en clase (creciente/decreciente, continua/discontinua, biyectiva/inyectiva/suprayectiva, trascendente/algebraica) la función  $f(x) = 3x^2 - 4$ .

Existe mucha disparidad entre las funciones intelectivo-analíticas y resolutivo-prácticas que requiere cada reactivo, por lo que no pueden, por ninguna circunstancia, mantener el mismo valor referencial para su apreciación juiciosa. Lo que nosotros hacemos y sugerimos es continuar con la misma dinámica anterior: primero esbozamos una estructura ponderativa basada en los rubros primarios, así es más sencillo apreciar la dimensión meritoria entera del examen. Cuando se ha decidido esto, los porcentajes se desglosan en cada rubro específico y así cada requisición queda ponderada conforme a un criterio más objetivo. Podrá parecer lo contrario, pero esta etapa es una de las más importantes y complicadas, y de ella depende gran parte de la ecuanimidad, objetividad y justicia de una evaluación. Es sumamente importante tener mucha claridad de lo que implica la atención y la resolución de cada reactivo para poder concederle el valor meritorio ponderativo correspondiente y justo.

Es muy importante también considerar otro factor dentro de esta etapa: el criterio de suficiencia/insuficiencia de aprobación. Cuando se ha estimado la ponderación porcentual para cada rubro se debe analizar, en conjunto, el criterio de satisfacción/insatisfacción de aprobación, donde la aprobación sea para quienes hayan conseguido evidenciar un nivel mínimo de suficiencia competencial considerado en la proporción ponderativa. Esto debe implicar la consideración explícita de los criterios de suficiencia sustentados en los objetos y desempeños mínimos que un alumno debió haber conseguido y lógicamente evidenciado; es decir, las funciones básicas examinatorias deben fundamentar los criterios de suficiencia/insuficiencia de un examen. Por ejemplo, si deseo examinar en trigonometría la clasificación angular y lateral, los procesos de perímetro, área y relación pitagórica en triángulos rectángulos, debo establecer los mínimos de suficiencia

para poder acreditar la examinación. En este caso, si el alumno domina los procesos de perímetro y área, pero no clasifica correctamente ni maneja los procesos de relación pitagórica (catetos, hipotenusa y el teorema de Pitágoras), no sería acreedor de la suficiencia, pues ambos conceptos son elementales pero no para la asignatura de matemáticas II de bachillerato. Cuando se establecen los porcentajes es necesario hacer algunas pruebas simples para encontrar que las combinaciones porcentuales blindarán los criterios de suficiencia/insuficiencia aprobatoria y garantizarán el mérito correspondiente.

### **Análisis de los niveles categóricos de dominio**

Luego de haber establecido las ponderaciones oportunas, se debe elaborar la subdivisión de los niveles recomendablemente posibles para cada rubro. Esto es, identificar cómo convenimos dividir cada una de las dimensiones de alcance de cada requisición en categorías de logro o niveles de dominio. Por ejemplo, una de las divisiones convencionales es: excelente, bueno, regular, suficiente e insuficiente. Esta clasificación categórica sirve para dimensionar cuál es la subdivisión conveniente para cada caso; todos los rubros pueden mantener diferentes subdivisiones, lo cual puede agilizar o entorpecer la construcción de la rúbrica. En el ejemplo de arriba, donde se requería el concepto de función, establecer los cinco niveles de dominio sería un abuso práctico; muchos alumnos podrían definirlo en menos de dos renglones y significaría demasiado esfuerzo –ya lo hemos hecho– empatar el concepto en una clasificación tan específica. En este caso, lo más recomendable sería tener tres puntos de referencia que cotejen si el concepto es aprobable sin reparo, aprobable con objeciones y no aprobable. Lo que no podría pasar con “tabula, grafica y clasifica en todas las categorías revisadas en clase (creciente/decreciente, continua/discontinua, biyectiva/inyectiva/suprayectiva, trascendente/algebraica) la función  $f(x) = 3x^2 - 4$ ”. Ahora revisemos esto con atención.

Primero habría que desglosar en rubros específicos el primario. Podría ser:

- a) tabulación de la función,



- b) graficación de la función,
- c) clasificación 1,
- d) clasificación 2,
- e) clasificación 3,
- f) clasificación 4.

Si tomamos de ejemplo el primer rubro específico, podríamos establecer una clasificación más adecuada al proceso de tabulación: a) aprobable sin reparo, b) aprobable con objeciones, c) suficiente para aprobar, y d) no aprobable. Esta consideración –cuádruple– permitiría tener una dimensión más específica del nivel de dominio de un rubro que lo amerita. La tabulación es un recurso sumamente útil en la revisión del tema de funciones, debe tener la atención suficiente y la valoración necesaria.

Lo que tiene que quedar muy en claro en esta etapa es lo siguiente: aquí debemos establecer la clasificación de los niveles convenientes para cada rubro, en materia concreta, el número útil de categorías y el esbozo de su dimensión de logro, ya que no es sino hasta la siguiente etapa cuando se definirá con detalle cada uno de los indicadores específicos. En realidad, esta etapa es bastante rentable tanto para la agilización posterior de la rúbrica como para objetivación de la evaluación.

### **Determinación de los indicadores**

Los indicadores son los constructos categóricos de logro de un rubro. Son los enunciados que indican el dominio cualitativo de un fragmento competencial, desempeño u objeto de aprendizaje específico. En la etapa anterior se define el número de indicadores que existirán y se bosqueja la naturaleza del indicador. En esta etapa el indicador se desarrolla con formalidad y se amplía la dimensión de su alcance. Para la referencia cualitativa estos indicadores son muy útiles, proporcionan un esquema basto del nivel de logro de cada rubro y permiten la apreciación más real y objetiva de cada desempeño.

La función primaria de los indicadores es orientar al evaluador-intérprete en la toma de decisiones estimativa. Cuando existen procesos, operaciones o desarrollos que valorar, la amplitud

de apreciación es muy considerable y resulta indispensable el apoyo de guías estandarizadas construidas con el fin de aproximar, con objetividad, las estimaciones verbales.

En la etapa anterior, se definen los niveles categóricos de dominio, delineando los indicadores tentativamente, como aprobable sin reparo, aprobable con objeciones y no aprobable, por ejemplo. Esta etapa debe confeccionar los indicadores de manera más abundante y específica, de tal modo que terminen figurando como constructos semánticos potencialmente indicativos. No significa que deban ser enunciados largos y abrumadores, sino frases concretas, claras y significativas.

Para el caso que hemos manejado, donde se requisita el concepto de función, si se acordó manejar tres niveles categóricos, los indicadores tendrían que ser mucho más específicos que simplemente aprobar sin o con objeciones. De este modo podrían ser:

- Nivel 1 (máximo): clara, completa y con los elementos suficientes.
- Nivel 2 (medio): nivel 1 con estándares aceptables.
- Nivel 3 (bajo o nulo): confusa, incompleta y sin suficientes elementos.

Es verdaderamente elemental que el evaluador-intérprete construya un concepto estimativo-relativo-cualitativo de los datos recogidos del instrumento parcial y completo en su revisión a partir de estos indicadores, pues de este devendrán las siguientes etapas y, por lo tanto, las reinterpretaciones escalares, los juicios y las decisiones.

### **Proporción y distribución de puntajes**

Durante esta etapa es necesario revalorar, de manera general, la dimensión ponderativa porcentual que se estableció con anterioridad, ya que los porcentajes asignados a cada rubro se deberán desglosar en los niveles de dominio.

Este trabajo puede realizarse de diversas formas, pero lo más importante es, por un lado, respetar los porcentajes designados



en cada rubro y, por otro, estimar con prudencia y equidad los puntajes correspondientes en cada nivel.

Una de las formas que nosotros trabajamos y recomendamos es la de la asignación de puntaje por rangos porcentuales máximos y mínimos. Si el rubro cubre el 10% del valor ponderativo, los puntajes se distribuyen en los niveles por segmentos de un valor mínimo a un valor máximo correspondientes a sus niveles superior e inferior. Por ejemplo, para el nivel “clara, completa y con los elementos suficientes”, sus valores podrían ser de 10% a 7%, para el de “nivel 1 con estándares aceptables”, de 6% a 3% y para el de “confusa, incompleta y sin suficientes elementos”, de 2% a 0%. Deberá notarse que la distribución se hace con rangos de cuatro valores para los dos niveles más altos y de tres para el más bajo; esto obedece a una razón importante. Los niveles más altos presentan mayor amplitud de apreciación, los rasgos calificativos de “claridad, entereza y la totalidad de los elementos” implican un más amplio sentido interpretativo y juicioso que la “confusión” o la “no entereza”. Aunque esto, en apariencia, pueda resultar trivial, cobra una importante connotación cuando se valora holísticamente el instrumento y se realiza el conteo del puntaje.

Una de las principales ventajas de este modelo de distribución es que solo es necesario contabilizar los porcentajes y dimensionar el aprovechamiento evidenciado en el instrumento para asignar la calificación, aunque no en todos los casos es el más recomendable.

Otra forma de distribuir el puntaje es asignando puntos que correspondan a los niveles de dominio de manera directa. Por ejemplo: para “clara, completa y con los elementos suficientes”, tres puntos; “nivel 1 con estándares aceptables”, dos puntos y para el de “confusa, incompleta y sin suficientes elementos”, un punto, y aquí podría asignarse, convenientemente, un nivel adicional para identificar el puntaje cero, que sería cuando “no existe definición” o cuando existe “sin coherencia verídica”. Para este modelo es necesaria la construcción de una escala que ensamble los puntos correspondientes con las ponderaciones porcentuales correspondientes y con la escala del 100% calificativa, o de otro modo balancear los puntajes para que en sí mismos contemplen ya la ponderación asignada.



Este es el último paso para la confección de la rúbrica. Es decir, de esta etapa debe emerger el instrumento cualitativa y cuantitativamente finalizado. Lo restante es evaluarlo.

## Piloteo

La fase de piloteo, como todo proceso de testeo, es la evaluación y la aprobación del instrumento. Pero, ¿cómo pilotear la rúbrica?

Nuestra experiencia nos mostró que, en las primeras ocasiones, es necesario contestar un examen y analizar en qué partes, aspectos y elementos pueden existir incongruencias o desproporciones: desbalances ponderativos, insuficiencias categóricas, discrepancias semánticas, etcétera.

Debemos reconocer que, al principio, esta etapa fue una tarea muy difícil, laboriosa, amplia y extenuante. Encontrábamos múltiples aspectos que conveníamos modificar del diseño inicial porque estábamos convencidos de que estaba “mal proporcionada o insuficientemente específica”. Esto habla de que pudimos percatarnos del trabajo que significa aproximarse a la objetividad, a la justicia y la calidad integral, así como encontrar factores que reducían nuestros márgenes de proximidad a la estandarización conveniente, que logramos establecer a lo largo de todas estas líneas anteriores.

Con algunos ejercicios de práctica, pudimos hacer el piloteo con exámenes reales de alumnos evaluados, sin que esto significara que esos exámenes se encontraran en desventaja de apreciación y valoración. Lo recomendable es elegir el examen de un alumno de resultados promedio, revisarlo cuidadosamente y valorar si el resultado arrojado converge con la apreciación cualitativa del aprovechamiento evidenciado en el instrumento, de tal manera que en él podamos ver el mérito aprobatorio/reprobatorio, así como su correspondiente simbólico escalar –recordemos que el mérito evidenciado debe determinar esta condición.

## REVISIÓN DEL EXAMEN MEDIANTE LA RÚBRICA

La fase de elaboración del instrumento verificador constituye uno de los puntos vertebrales de la interpretación del examen. El otro



es el de la implementación de este. Al igual que la confección de la rúbrica, su uso se puede dividir en etapas que ayudan a reconstruir los procesos, paso a paso, a fin de explicitar con detalle cada parte como un subsistema.

Para verificar un examen mediante una rúbrica es necesaria la progresión de las siguientes etapas.

### **Lectura y decodificación de datos**

La primera etapa de la verificación consiste en realizar una lectura asertiva de la información recabada en relación con las requisiciones de los ítems. Esto, de algún modo, ya está digerido en los rubros, así que lo que resta es leer con atención lo acopiado e interpretarlo dentro del contexto socioacadémico. Esto significa que las respuestas presentan una carga semiótica correspondiente al grupo cultural, edad y sesgo académico que puede diferir de la información esperada, puramente académica. Por ejemplo, si un ítem solicita la clasificación de los triángulos por la medida de sus lados de forma textual y nutrida, no deberíamos ser tan estrictos para esperar una respuesta con un lenguaje netamente científico como “la razón categórica que existe entre las dimensiones de los lados de un triángulo se manifiesta como una tipificación tripartita que considera: a) equiláteros, a aquellos que presentan lados equilongitudinales, b) isósceles, a aquellos que presentan dos lados equilongitudinales, y c) escalenos, a aquellos que presentan longitudes distintas en sus lados”; por el contrario, aunque es efectivamente deseado el uso apropiado del lenguaje matemático, puede ser muy productivo y valioso que el alumno se exprese con propiedad dentro de sus limitantes comunicativo-culturales y así se le valore; por ejemplo, “está el equilátero, que tiene sus tres lados iguales, luego el isósceles, que tiene dos lados iguales y otro no, y el escaleno que tiene tres lados diferentes”. Aquí es necesario hacer hincapié en que resulta indispensable leer con atención, pero además con un sentido contextual, la información recabada en cada ítem para extraer su sustrato esencial y verificarlo con la rúbrica. De este modo, no es solo leer, sino decodificar, contextualizar y comprender.

Expongamos un ejemplo real. La respuesta de un alumno al ítem “define lo que es una función y una relación” (en el caso solo de la función), fue: “la función se define como una dependencia de una cosa a otra”. Otro alumno: “una función es una acción que se ejerce en una relación, puede ser entre dos personas en la cual una puede depender de la otra”. Otro: “es la relación que hay entre dos individuos y que los dos dependen entre sí”. Analicemos lo matemáticamente correcto de cada respuesta. Leibniz se refería con una función a la relación de dependencia de dos o más variables o cantidades, un concepto que se revisó en clase. Si nos basamos en este para verificar las respuestas, la primera define con gran amplitud el concepto vertebral: dependencia. En efecto, no menciona que es una relación y califica como “cosas” los objetos matemáticos (variables o cantidades), pero podríamos considerar que el concepto que logró construir tiene suficiente coherencia semántica y demuestra haber adquirido la parte nuclear de la definición, con alguna penalización por no haber referido el término “relación”; es decir, no tendría el máximo nivel categórico de dominio. En el segundo caso podemos notar que el concepto necesita organización semántica pero es aceptable y además incluye las categorías “relación” y “dependencia”; sin embargo, la correlación que hace deja ver que el concepto quedó matemáticamente descontextualizado, pues señala “dos personas”, no dos objetos matemáticos, y aunque lo indica como una analogía, “puede ser”, que sugiere que es un ejemplo donde la función puede darse no limitándola a ese contexto, es evidente que el concepto necesita mayor especificidad; no obstante, es suficientemente aceptable con alguna penalización por ello. En el tercer caso tenemos algo similar. El concepto posee las palabras nucleares, es semánticamente incorrecto, lo relaciona a dos sujetos –y esto se debe a que en clase se ejemplificaron a personas para indicar una relación de dependencia–; sin embargo, no lo maneja como una analogía como en el caso anterior, sino que lo limita al hecho de “la relación que hay entre dos individuos”. Pero hay una connotación muy importante que genera otras condiciones evaluativas: “los dos dependen entre sí”. Aunque no es sencillo de interpretar, lo claro es que la dependencia de una función se señala como recíproca, lo cual es matemáticamente incorrecto.



Esto indica, entonces, que el concepto elaborado es disfuncional y tendría que ser penalizado con rigor, es decir, encuadrarlo en el rubro correspondiente donde se señale que el concepto es disfuncional, con elementos nucleares que podría ponderarse en un 50% del valor nominal del ítem por tener los dos conceptos nucleares en una redacción pobre y con disfunciones con respecto a la definición de Leibniz.

La lectura y la decodificación de las respuestas de los ítems deben ofrecer al evaluador-intérprete una dimensión clara de la evidencia recopilada, no solo en términos de resolución-satisfacción sino como una manifestación cognitiva, consecuencia de un conflicto intelectual.

Muchas veces las respuestas de los alumnos son devaluadas por los maestros sin fundamento. La pobreza gramatical, las deficiencias grafológicas o el mensaje semántico confuso de una respuesta son motivos por los que evaluadores pueden discriminar o anular el valor ponderativo de un producto recabado. El modelo de evaluación Valoración de competencias matemáticas basado en la examinación e interpretación cuantitativo-cualitativa (VCM-EICC) (Mejía, 2013) en el que a su vez está basado este estudio, sublima la recuperación fiel, contextual y racional de las respuestas de los ítems como el núcleo de la interpretación-valoración dentro del complejo evaluativo.

### **Análisis cualitativo y estimación de logros**

La información recuperada debe someterse a un proceso juicioso en primer término cualitativo, que permita una apreciación holística de los progresos conseguidos por el sujeto evaluado y que lo comience a situar en una perspectiva adjetiva. Con ello nos referimos, como hemos ido insistiendo en el método interpretativo cualitativo-cuantitativo, a que el evaluador requiere dimensionar el aprovechamiento examinatorio en relación con los productos evidentes decodificados, como una serie de resultados adjetivos que converjan en la matriz de verificación. El esmero por elaborar una rúbrica se debe visualizar en esta etapa de la revisión. Puesto que la primera fase interpretativa es cualitativa, el evaluador necesita elaborar un esquema mental de los logros demostrados y va-

lorar, adjetivamente, sus niveles de dominio con amplitud verbal; esto es muy funcional cuando confrontamos nuestros criterios y juicios adjetivos con la rúbrica, porque nos costará menos trabajo empatar nuestro juicio con los niveles establecidos.

Esta etapa tiene dos funciones: hacer coincidir nuestra apreciación en un nivel de dominio –en cada rubro– e ir construyendo, a lo largo de la revisión del instrumento examinador, una valoración completa e integral para apreciar el logro final y sustentar la suficiencia/insuficiencia de aprobación.

### Encuadre cuantitativo

Luego de la estimación adjetiva, es decir, de valorar cualitativamente la información acopiada, se procede a la conversión escalar. Podrá advertirse que este proceso puede realizarse de distintas maneras de acuerdo con el modelo de distribución de puntajes. Si se eligió alguno que vaya sumando puntos dentro de una escala distinta al 100%, en esta etapa se realiza la escala adecuada y se convierte, de manera proporcional, a un número inteligible dentro de la escala –porcentualmente– oficial. Se revalora el resultado sumado y se establece una tolerancia de error tasado, que significa una posibilidad de reencuadrar el resultado numérico conforme a una relación más estrecha con la apreciación adjetiva global. Generalmente, esta tolerancia es una oscilación entre los rangos escalares de los porcentajes oficiales; esto es, si el máximo de puntos es de 25, equivaldría al 100% en la escala calificativa, 24 puntos equivaldría a 96% y 23 puntos a 92%; la tolerancia de error sería de 6%, tres puntos hacia arriba y tres puntos hacia abajo, donde los 24 puntos podrían convertirse en 99% o 93% en relación con la apreciación adjetiva, permitiendo así la movilidad del evaluador-intérprete en un rango de reencuadre.

Para una distribución directamente porcentual –en la escala oficial–, el encuadre puede ser un proceso alternativo a la estimación de logros cualitativa. Cuando el evaluador encuadra su juicio en un nivel de dominio, en ese momento puede asignar un porcentaje directo del rango establecido –donde está ya contemplada la tolerancia de error–. Aunque también es posible no hacerlo de manera alterna, sí lo recomendamos. Cuando solo se estima

adjetivamente y no se cuadra en un porcentaje, el regresar a hacerlo implica más tiempo, pues el evaluador debe recapitular la situación adjetiva apreciada de nuevo para estimar su porcentaje. Este proceso de encuadre también permite la consideración de las posibilidades de plagio durante la examinación. Si un alumno presenta síntomas de plagio, la tolerancia de error podrá apoyar a revalorar los resultados bajo ese precepto.

En esta etapa, es muy importante hacer un cotejo y una revaloración meritoria del resultado escalar. Algunos profesores llegan a asombrarse de los resultados arrojados en sus exámenes; esto se debe a que sus instrumentos no proporcionan ponderaciones justas y los resultados se disparan sin correspondencia meritoria.

## CONCLUSIONES

La examinación sistémica es un proceso que intenta, con muchas oportunidades de logro, recuperar información relevante y fidedigna de la realidad escolar de los alumnos, para ser procesada de manera sensata y objetiva y apoyar a los docentes, significativamente, para construir sus juicios y calificaciones; esto en aras de ofertar a los alumnos un sistema verídico y justo de su propio aprovechamiento escolar y su desempeño académico en matemáticas, además de tener a la mano elementos notables para ir reconfigurando su propia intervención didáctica y mejorar su participación como docente y como evaluador.

La implementación de la examinación sistémica no es sencilla. Efectivamente, es una serie de procesos que merece consagración —tiempo, dedicación, análisis, conflicto—. Sin embargo, es un recurso muy valioso para el quehacer docente de aquellos quienes participamos activamente en la mejora continua de la educación en México y Latinoamérica; en este caso, en la preocupación de fomentar una evaluación más justa, ordenada y útil en las aulas de asignaturas matemáticas.



## REFERENCIAS BIBLIOGRÁFICAS

- Barajas**, Rosa María y María del Carmen Gilio. “El examen dentro de un proceso de evaluación participativa”, en *X Congreso Nacional de Investigación Educativa*, Veracruz, 25 de septiembre de 2009. Disponible en [http://www.comie.org.mx/congreso/memoriaelectronica/v10/pdf/area\\_tematica\\_14/ponencias/1339-F.pdf](http://www.comie.org.mx/congreso/memoriaelectronica/v10/pdf/area_tematica_14/ponencias/1339-F.pdf)
- Bertely**, María. *Conociendo nuestras escuelas: un acercamiento etnográfico a la cultura escolar*, México, Paidós, 2000.
- Cámara de Diputados del H. Congreso de la Unión**. “Ley General de Educación”, en *Diario Oficial de la Federación*, 13 de julio de 1993. Disponible en <http://www.diputados.gob.mx/LeyesBiblio/pdf/137.pdf>
- Castañeda**, Sandra (coord.). *Evaluación del aprendizaje en el nivel universitario: elaboración de exámenes y reactivos objetivos*, México, UNAM/Conacyt, 2008.
- Díaz Barriga**, Ángel. *El examen: textos para su historia y debate*, México, CESU-UNAM, 1993.
- García**, Juan Antonio. “La didáctica de las matemáticas: una visión general”, en *Matemáticas en Secundaria*, 2001. Disponible en <http://www.gobiernodecanarias.org/educacion/rtee/didmat.htm>
- López**, Miguel. “Preparación y calibración de reactivos: medición de y para el aprendizaje”, en Sandra Castañeda (coord.). *Evaluación del aprendizaje en el nivel universitario: elaboración de exámenes y reactivos objetivos*, México, UNAM/Conacyt, 2008.
- Medina-Díaz**, María del R. y Ada Verdejo-Carrión. *Evaluación del aprendizaje estudiantil*, San Juan, Isla Negra, 2001.
- Mejía**, Omar. “De la evaluación tradicional a una nueva evaluación basada en competencias”, en *Revista Electrónica Educare*, 16 (1), abril de 2012, pp. 27-46. Disponible en <http://www.revistas.una.ac.cr/index.php/EDUCARE/article/view/3737>
- Mejía**, Omar. “Hacia un modelo de evaluación matemática en bachillerato: VCM-EICC, Valoración de competencias matemáticas basada en la examinación e interpretación cuan-



titativo-cualitativa”, ponencia presentada en el II Congreso de investigación de matemáticas en ciencias sociales y educación superior aplicaciones y problemas de enseñanza-aprendizaje, Ciudad de México, junio de 2013.

**Rodríguez**, Héctor M. y Enrique García. *Evaluación en el aula*, México, Trillas, 1990.

**Santos**, Miguel Ángel. *Hacer visible lo cotidiano. Teoría y práctica de la evaluación cualitativa de los centros escolares*, Madrid, Akal, 2008.

