Una crítica de las calificaciones, los sistemas de calificación y las prácticas de calificación

Lorin W. Anderson
Distinguido profesor emérito de Carolina
Universidad de Carolina del Sur, Columbia (EE.UU)
1 de septiembre de 2016

“Cuando tomamos en cuenta el uso casi universal de un sistema de notas en todas las instituciones educativas, no podemos menos que asombrarnos ante la fe ciega que se ha tenido en la confiabilidad de los sistemas de notas”

Finkelstein, 1913

“¿Acaso no es una hipocresía predicar sobre la importancia de la innovación en la educación mientras nos aferramos a un sistema de calificación casi tan arcaico como inútil?”

Ferriter, 2015

Cinco preguntas básicas

1. ¿Por qué calificamos a los alumnos?
2. ¿Qué significan las calificaciones?
3. ¿Qué tan confiables son las calificaciones de los alumnos?
4. ¿Qué validez tienen las calificaciones de los alumnos?
5. ¿Qué consecuencias tienen las calificaciones para los alumnos?

Mi intento

El *Merriam Webster Learner's Dictionary* define la critica como un "juicio ponderado en el que se da la opinón sobre lo bueno y lo malo de algo".

Algunas definiciones

- UNA CALIFICACIÓN es una posición en un continuo de rango, calidad, competencia, intensidad, o valor. El continuo indica gradaciones de rango, calidad, y así por el estilo. Estas gradaciones se pueden expresar numéricamente (e.g., 1 a 100), con letras (por ejemplo, A, B, C, D, F), o usando un conjunto de descriptores verbales (verbigracia, ejemplar, competente, elemental, menos que elemental).
- CALIFICAR a un alumno significa colocarlo en el continuo con base en impresiones, evidencias o ambas cosas.
- *[Nota. A principios del siglo XX, el término "nota" se usaba en lugar de "calificación". Los académicos y practicantes hablaban de "sistemas de notas" en lugar de "sistemas de calificación". A los alumnos les otorgaban "notas" en lugar de "calificaciones".]*

¿Por qué calificamos a los alumnos?

- Para motivarlos
  - "En el mejor de los casos, las calificaciones motivan a los estudiantes de alto aprovechamiento a seguir obteniendo buenas calificaciones, independientemente de que esa meta coincida por casualidad con el aprendizaje. En el peor, las calificaciones disminuyen el interés por aprender y acentúan la ansiedad y la motivación extrínseca, especialmente en los estudiantes que tienen dificultades(Schinske & Tanner, 2014).
- Para comunicarnos con una diversidad de públicos
  - Los públicos incluyen a padres de familia, administradores escolares, maestros y administradores de educación de la secundaria, empleadores, los que elaboran políticas (incluyendo a legisladores) e integrantes de los medios.
  - Los distintos públicos requieren formas diferentes para exponer y explicar las calificaciones.
- Para dar retroalimentación a los maestros

Parte de una boleta de calificaciones basada en estándares

NOMBRE DE LA ALUMNA: MARIA D. IBARROLA	CALIFIC.
Usa el equipo de laboratorio de manera adecuada y segura	4
Calcula correctamente la densidad	4
Aplica el concepto de densidad a problems relevantes	2
Recuerda las fórmulas de las leyes de los gases (e.g., Boyle, Gay-Lussac)	4
Elige la ley apropiada de los gases para resolver problemas específicos	1

Nota. Las calificaciones son: excelente (4), competente (3), se acerca a competente (2), muy por debajo de competente (1).

¿Qué significan las calificaciones?

- "Me incliné sobre el hombro del alumno… y le pregunté si me podía mostrar la retroalimentación de su maestra sobre su trabajo y sus notas actuales. Abrió su carpeta electrónica de ciencias sociales en su laptop y había una lista de tareas. … Al lado de una de las tareas decía 100%. Le pregunté qué significaba eso: 'Pues, esa la entregué a tiempo', contestó" (Tinney, 2014, p. 1).
- Para determinar el significado de las calificaciones, se deben contestar dos preguntas fundamentales.
  - ¿Qué representa la calificación?
  - ¿Cómo se interpreta la calificación?

¿Qué representa la calificación?

- Una calificación es más significativa cuando representa una sola variable bien definida (por ejemplo, aprovechamiento en ciencias, esfuerzo invertido, conducta social positiva).
  - Individual significa que una calificación no es una combinación indefinida de aprovechamiento, esfuerzo, conducta y otros factores diversos.
  - Bien definida significa que le queda claro a todo el mundo lo que "cuenta" para la calificación. ¿Una calificación de aprovechamiento representa memorización, pensamiento crítico, o ambos? Si ambos, ¿qué porción de la calificación representa cada uno? ¿Qué evidencias de esfuerzo (o su falta) toma en cuenta el maestro cuando otorga una calificación de "esfuerzo"?

¿Cómo se interpretan las calificaciones?

- Durante las primeras décadas del siglo XX, se popularizó el "calificar sobre la curva". Con esta práctica, un cierto porcentaje de alumnos recbía A, B, C, D, y F. En 1914, con base en una serie de estudios técnicos, se recomendó una distribución de calificaciones de 7% A, 24% B, 38% C, 24% D, y 7% F. La calificación de un alumno, entonces, indicaba dónde se ubicaba con relación a los demás.
- En 1963, Robert Glaser sugirió que los resultados de los exámenes (y, por asoción, las calificaciones) debían interpretarse en términos de un continuo de aprovechamiento subyacente que abarcara desde nada competente hasta completamente competente. La calificación de un alumno indicaba su lugar en el continuo subyacente.
- *[Nota. Una calificación de aprovechamiento que represente el grado de dominio que el alumno ha alcanzado con respecto a conceptos y habilidades clave a lo largo de un continuo también se puede interpretar hacienda referencia a un norma. No obstante, lo contrario no es cierto. [Repasar la Diapositiva 6.]]*

Reencuentro con el significado

- "'Cuando uso una palabra', dijo Humpty Dumpty, con tono algo despreciativo, 'significa justamente lo que yo quiera que signifique: ni más, ni menos.' 'La pregunta es,'" dijo Alicia, 'si es posible lograr que las palabras signifiquen tantas cosas distintas'" (Lewis Carroll, A través del espejo, 1871). Cuando se trata del significado de las calificaciones, la respuesta a la pregunta de Alicia es: "¡Si!"
- "Parece muy claro que ni el número ni la letra han transmitido un significado común al alumno, a la maestra, y al administrador por igual. No debemos olvidar que estos tres grupos tienen que 'pensar juntos' con respecto a las notas escolares, y nuestro propósito es elaborar un sistema de notas sencillo y económico en su aplicación, y, sin embargo, uno que permita a estos tres conjuntos de mentes estar de acuerdo con las notas que se pondrán a los resultados del aprendizaje" (Rugg, 1918, p. 713).

Un breve interludio

- Antes de seguir, debo hacer una importante distinción entre una calificación *individual* (es decir, una calificación que se otorga a un trabajo individual de un alumno) y una calificación acumulativa (es decir, una calificación que se otorga a un cúmulo de trabajos de un alumno). Con respecto a las calificaciones acumulativas, cada muestra de trabajo dentro del cúmulo se califica individualmente; sin embargo, el cúmulo como un todo también recibe una calificación.
- Las calificaciones acumulativas requieren de algún tipo de recoleción de datos, ya sea un simple promedio aritmético de las calificaciones individuales, un simple promedio aritmético después de eliminar la calificación más baja y la más alta, un promedio ponderado (como cuando un examen de una unidad cuenta el doble que las tareas en casa), o algún otro método. Las calificaciones acumulativas también se pueden producir sumando diferentes materias impartidas por distintas maestras para producir un promedio general (PG).

¿Qué tan confiables son las calificaciones individuales ?

- Con respecto a las "calificaciones individuales", las respuesta es muy sencilla: "¡Son muy poco confiables!" Starch (1913) sugirió que había cinco posibles fuentes de la poco confiabilidad:
  - (1) diferencias provocadas por la incapacidad de los maestros para "distinguir entre grados estrechamente relacionados de méritos" (p. 630),
  - (2) diferencias en los criterios empleados por diferentes maestros (por ejemplo, contenido, mecánica, estilo de calificación en ensayos y trabajos escritos),
  - (3) diferencias en los estándares de calidad usados por diferentes maestros (por ejemplo, ¿que constituye la excelencia?),
  - (4) diferencias entre los estándares de distintas escuelas, y
  - (5) diferencias en la forma en que los maestros distribuyen sus calificaciones entre los alumnos.

Fuentes y soluciones

Fuente	Soluciones propuestas históricamente
Inhabilidad para diferenciar entre puntos porcentuales	Cambiar de calificaciones de porcentajes a letras "Los maestros pueden manejar cinco divisiones".
Uso de criterios diferentes	Guías estandarizadas de calificación (una de las primeras se desarrolló en 1915)
Uso de estándares de calidad diferentes	Cálculo y uso del "factor de correción" del maestro basado en si el maestro era "barco" o "duro" para calificar
Distribuciones de calificaciones diferentes	Por ciento estandarizado de As, Bs, Cs, Ds, y Fs basado en una supuesta "óptimar" distribución subyacente de calificaciones.
Criterios/estándres diferentes en escuelas distintas	No se propusieron soluciones; estas diferencias siguen siendo problemáticas hoy en día.

¿Mejoran las guías la confiabilidad de las calificaciones individuales?

- Estudio de Starch y Elliott (1912)
  - Materia = inglés
  - Se les pidió a maestros de inglés de preparatoria calificar los mismo dos ensayos usando una escala de 100 puntos
  - En ambos ensayos, las calificaciones oscilaron entre 50 y 98
- Estudio de Brimi (2011)
  - Materia = inglés
  - Se les dio a maestros de inglés de preparatoria siete días de capacitación (una semana en el verano y un fin de semana durante el año escolar) sobre el uso de una guía de redacción para calificar ensayos
  - Después de terminar la capacitación, se les pidió a los maestros calificar el mismo ensayo en una escala de 100 puntos
  - Las calificaciones oscilaron entre 50 y 96.

¿Qué tan grande es el problema de las diferencias entre escuelas?

- En 1994, el Departamento de Educación de Estados Unidos comparó los resultados de pruebas nacionales (NELS:88) y las calificaciones en escuelas de mucha pobreza y más prósperas, según declaraciones de los mismos alumnos.
- Los alumnos en escuelas de much pobreza que recibieron más que nada A en inglés lograron casi las mismas puntuaciones en pruebas de lectura que los alumnos de C y D en las escuelas más prósperas.
- Los alumnos de escuelas de mucha pobreza que recibieron más que nada A en matemáticas tenían más o menos las mismas puntuaciones en pruebas de matemáticas que los alumnos de D en las escuelas más prósperas.
- *[Nota. Estoy consciente de que este ejemplo depende de calificaciones acumulativas más que de calificaciones individuales, pero encaja mejor aquí que después.]*

¿Qué tan confiables son las calificaciones acumulativas?

- En 1914 T. L. Kelley examinó la relación entre las calificaciones que los alumnos recibieron en los grados 6° and 9°. El coeficiente de correlación fue de 0.73.
- En 1964, B. S. Bloom resumió una serie de estudios que investigaron la estabilidad de las calificaciones de alumnos a lo largo de periodos de tres a ocho años. Los coeficientes de relación oscilaron entre 0.70 a 0.85.
- Los resultados de varios estudios recientes realizados a nivel universitario sugieren que la confiabilidad de los promedios generales de calificación es bastante alta (con coeficientes alfa que va de 0.72 a 0.91 en el transcurso de dos, cuatro, seis y ocho semestres).
- En general, por lo tanto, las calificaciones acumulativas son bastante estables al paso del tiempo (con aproximadamente la mitad de la varianza en las calificaciones otorgadas en algún punto del tiempo explicadas por la varianza en las calificaciones otorgadas en algún punto del tiempo anterior.

¿Qué explica la diferencia en la confiabilidad de las calificaciones individuales y las acumulativasof?

- Existen por lo menos dos posibles explicaciones.
  - Estamos tratando con dos definiciones muy distintas de confiabilidad. Para las calificaciones individuales, el énfasis está en el acuerdo entre quienes califican. Para las calificaciónes acumulativas, el énfasis está en la estabilidad al paso del tiempo.
  - Prácticamente en cada caso, la recolección de datos aumenta la confiabilidad. Así como los exámenes más extensoss son más confiables que las cortos (considérese, por ejemplo, una prueba de un solo elemento), así también los promedios de calificaciones que se basan en ocho semestres son más confiables que los que se basan en dos.
- Lo importante aquí es que el grado de confiabiidad de las calificaciones depende de (1) la definición de confiabilidad y (2) si estamos tratando con calificaciones individuales o acumulativas. Las personsas que hacen declaraciones tales como "las calificaciones no son confiables" están desinformadas.

Disposición de datos para estudios de confiabilidad temprana

Alumno hipotético	Mtro 1	Mtro 2	Mtro 3	Mrto 4	Mtro 5	Total
Número 1	80	60	30	70	90	66

Data Layout for GPA Reliability Studies

Identificación de Alumno	1er sem	2º sem	3er sem	4º sem	5º sen	6º sem	7º	8º
0001	1.5	1.0	1.5	2.5	3.0	1.5	2.0	3.0
0002	2.0	2.0	3.0	2.0	1.5	1.5	3.0	3.0
0003	2.0	1.5	2.0	3.0	1.5	3.0	1.5	3.0
0004	2.5	2.0	2.5	2.5	2.0	2.0	3.0	3.0
0005	2.5	3.0	3.0	2.0	2.5	2.5	1.5	3.0
0006	2.5	2.5	2.5	3.0	2.0	3.0	3.0	4.0
0007	3.0	2.0	3.0	3.0	2.5	2.5	3.0	4.0
0008	3.0	3.0	4.0	2.5	3.0	4.0	3.5	4.0
0009	3.0	2.5	3.0	4.0	3.5	3.5	4.0	4.0
0010	3.0	3.0	3.5	3.5	3.5	3.5	4.0	4.0

Disposición de datos para estudio "expandido" de confiabilidad temprana

Alumnos hipotéticos	Mtro 1	Mtro 2	Mtro 3	Mtro 4	Mtro 5	Total
Número 1	80	60	30	70	90	66
Número 2	70	40	10	50	60	46

¿Qué tan válidas son las calificaciones?

- La respuesta a esta pregunta requiere respuestas a dos preguntas:
  - ¿En qué medida obtienen mejores calificaciones los alumnos que aprenden más?
  - ¿Hasta qué punto predicen las calificaciones el éxito subsecuente en la escuela o en la vida?

¿En qué medida logran mejores calificaciones individuales los alumnos que aprenden más?

- Es muy difícil responder a esta pregunta directamente. Basándonos en lo que se conoce en general, podemos decir que la validez de las calificaciones individuales mejora cuando
  - (1) la calificación representa un, y sólo un, atributo, en este caso aprovechamiento académico,
  - (2) existe un alto grado de alineación entre el plan de estudios, la enseñanza, y la evaluación, y
  - (3) los profesores son capaces de determinar con bastante precisión grados de aprendizaje o aprovechamiento académico.

¿En que medida logran mejores calificaciones acumulativas los alumnos que aprenden más?

- Para contestar esta pregunta empíricamente, debemos suponer que las pruebas que se aplican a los alumnos son indicadores válidos de su aprendizaje. Si aceptamos esta premisa, luego podemos inferir que los alumnos con puntuaciones más altas en las pruebas aprenden más. Si estos alumnos logran también mejores calificaciones, luego podemos concluir que los alumnos que aprenden más obtienen mejores calificaciones.
- Con base en estudios realizado a lo largo de los últimos 50 años, existe una correlación estadísticamente signficativa entre los resultados de pruebas estandarizadas y las calificaciones. La magnitud de los coeficiente tiende a fluctuar entre 0.5 y 0.6.
- En los últimos 25 años, los exámenes de fin de curso se han diseñado e implementado en muchos estados. En virtud de su relación con estándares académicos adoptados por los estados, es probable que estos exámenes se encuentren entre las mediciones más válidas que tengams con respecto al aprendizaje de los alumnos. Existen considerables pruebas de que los alumnos que aprueben los exámenes de fin de curso también logran calificaciones más altas en el curso. Una vez más, por lo tanto, existe cierta evidencia de que los alumnos que aprenden más obtienen mejores calificaciones.

¿Hasta que punto las calificaciones acumulativas predicen el éxito subsecuente en la escuela o en la vida?

- Aunque los resultados difieren entre las universidades, por lo común el promedio general de la preparatoria represent entre 20 y 25 por ciento de la varianza en el promedio general en la universidad, tanto de los alumnos de segundo año como los de cuarto. Es el mejor predictor individual que tenemos del promedio general en la universidad, donde los resultados de las pruebas SAT representan un siete por ciento adicional de la varianza.
- Los resultados de un estudio longitudinal de los alumnos de preparatoria con mejores calificaciones sugirieron que siguieron teniendo éxito en la universidad, con un promedio general de 3.6. La mayoría terminó por trabajar en profesiones tales como contabilidad, medicina, derecho, ingeniería, y educación.

¿Cuáles son las consecuencias de las calificaciones?

- Alfie Kohn ofrece una letanía de las consecuencias negativas de las calificaciones.
  - Tienden a reducir el interés de los alumnos en el aprendizaje mismo.
  - Tienden a reducir la preferencia de los alumnos por las tareas exigentes.
  - Tienden a reducir la calidad del pensamiento de los alumnos.
  - Dstorsionan el plan de estudios.
  - Desperdician mucho tiempo que podría dedicarse al aprendizaje.
  - Fomentan el hacer trampas.
  - Echan a perder las relaciones de los maestros con los alumnos.
  - Echan a perder las relaciones de los alumnos entre sí.

¿Qué pasaría se reemplazáramos "calificaciones" en las afirmaciones de Kohn con otras frases?

- [Los maestros, actividades y tareas aburridas] reducen el interés de los alumnos en el aprendizaje mismo (Baurelein, 2013).
- [Los mandatos y las regulaciones federales] distorsionan el plan de estudios (Robelen, 2011).
- [La conducta negativa de los maestros] echa a perder las relaciones de los maestros con los alumnos (Banfield, Richmond, & McCroskey, 2006).
- [El orden jerárquico y las camarillas] echan a perder las relaciones de los alumnos entre sí (McFarland, Moody, Diehl, Smith, y Thomas, 2014).

Calificaciones individuales vs. acumulativas: reprobar vs. fracasar

- Hace cuarenta años escribí lo siguiente: "El verbo 'fracasar' se refiere a la incapacidad de un individuo para lograr el éxito con respecto a una meta en particular. 'Fracasado' es un sustantivo que se refiere a una persona que, al no haber alcanzado una serie de metas relacionadas, se percibe como incapaz de obtener éxito en el futuro. … Fracasar es (o puede ser) benéfico para los individuos, mientras que ser un fracasado es casi siempre perjudicial".
- Obtener una sola calificación baja (incluso una F) tiende a no tener consecuencias graves. Obtener consistentemente calificaciones bajas (D y F) puede transformar "fracasar" en "fracasado".

Un estudio casi longitudinal

- Se incluyeron en el estudio dos grupos de alumnos en cada uno de cuatro grados (2, 4, 6, y 8). El Grupo A consistía de alumnos que habían estado en el 20 por ciento superior de la distribución de calificaciones en cada ciclo escolar. El Grupo B consistía de alumnos que habían ocupado el 20 por ciento más bajo.
- En el 2º grado, los dos grupos no diferían en términos del concepto académico que tenían de sí mismos. Para el 4º grado, empezaron a aparecer las diferencias en el concepto académico de sí mismos. Para el 8º grado, las diferencias estaban plenamente desarrolladas.
- Mi interpretación. Para el 2º grado, el grupo de abajo había fracasado sin duda, pero no se veían como fracasados. Para el 8º grado, la transición de fracasar a fracasado era completa.

Cinco conclusiones/recomendaciones

1. Se ha avanzado muy poco en los sistemas y las prácticas de calificación en el transcurso de los últimos 100 años

- "La educación masiva, con masas mucho más grandes mañana que ayer, requiere respuestas a estos problemas (de calificación). Y, sin embargo, hay poco interés. … Hay que elevar la calificación de su estatus actual como sólo una tarea más a su verdadera función … la evaluación del aprovechamiento de los alumnos y de la eficiencia de nuestras instituciones educativas" (Cureton, 1971, p. 8)
- Iba a usar una cita de un artículo escrito en 1902 sobre la clasificación del ganado, pero mi esposa lo consideraba apenas tangencialmente relacionado con mi propósito. Mi propósito era, sencillamente, que se ha avanzado más en los últimos 100 años en la clasificacion del ganado que en la calificación de estudiantes.

2. Para hablar de maner inteligente sobre las calificaciones es imperativo que diferenciemos las calificaciones individuales de las acumulativas

- "En resumen, lo que ofrecen las calificaciones es una precisión espuria: una apreciación subjetiva que se disfraza de una evaluación objetiva" (Kohn, 1999, p. 40). Ésta es una afirmación razonablemente acertada sobre las calificaciones individuales.
- "Advertimos que el aprovechamiento académico en la secundaria (definidos por el promedio general del 7° grado) es consistentemente un predictor importante de la terminación de la preparatoria, tomando en cuenta una variedad de otras características." (Kurlaender, Reardon, & Jackson, 2008, p. 17). Ésta es una afirmación razonablemente acertada sobre las calificaciones acumulativas.

3. Tenemos que diseñar & implementar sistemas de calificaciones que informen de manera efectiva a una variedad de públicos en formas que satisfagan las necesidades de información de esos públicos.

- "Cuidado con las dicotomías falsas" (Rich Shavelson, 9/2/2016).
  - Si cambiamos las etiquetas de "A" a "Excelente" o "Avanzado" y de "B" a "Muy por debajo de competente", ¿acaso hemos marcado alguna diferencia?
  - En muchos casos, proporcionar calificaciones basadas tanto en normas como en estándares tal vez sea conveniente y útil. Aunque la Facultad de Derecho de la Universidad de Wisconsin pone énfasis en la importancia de otorgar calificaciones sobre la base de cuánto o lo bien que han aprendido sus estudiantes, también declara muy claramente que "los estudiantes tienen un interés legítimo en conocer cómo se compara su desempeño académico con el trabajo de los demás".
- En lugar de abogar por un enfoque de calificaciones por encima de otro, como han hecho numerosos educadores (y lo siguen haciendo), la meta debe ser lograr un correspondencia entre el sistema de calificaciones, el público y las necesidades de información de ese público.

4. Tenemos que integrar plenamente la preocupación por las calificaciones en nuestras discusiones sobre el mejoramiento y la excelencia educativos.

- "En el transcurso de los últimos diez años se ha hecho cada vez más evidente que una de las causas que contribuyen al 'fracaso' de las escuelas públicas ha sido una mala administración del sistema de calificaciones. Reconociendo que parte de la dificultad quizá tenga su origen en la capacidad del alumno, que parte puede atribuirse claramente a un curso de estudio mal ajustado y parte a una docencia ineficiente, de todas maneras queda [para] la esmerada consideración de los administradores y los maestros [la necesidad de] reconstruir y administrar mejor el sistema de calificaciones (Rugg, 1918, p. 701)

5. Existe una falta descomunal de investigación sobre las calificaciones, los sistemas de calificación, y las prácticas de calificación.

- De hecho, nos estamos obligando unos a otros a aceptar toda suerte de vagos acuerdos sólo porque nadie cuenta con los hechos. ¿Quién sabe… si estamos siguiendo servilmente las tradiciones o luchando por un bien verdadero? ¿Quién sabe si los conservadores o los radicales tengan la razón? Lo que es más, ¿quién puede saberlo bajo las condiciones existentes? En lo personal, no estoy a favor de todas las tradiciones que se mantienen categóricamente, pero quiero afirmar con el mismo énfasis que no estoy a favor de adoptar sugerencias radicales sólo porque se ofrecen de manera persistente (Judd, 1910, p. 466).

¿Qué he aprendido al escribir esta presentación?

- “Nada es realmente bueno o malo por sí mismo: todo tiene que ver con lo que una persona piensa de ello ” (*Hamlet,* Acto 2, Escena 2)
- Necesitamos abogar menos por una postura y menos artículos en las páginas editoriales. Necesitamos más pruebas empíricas y diálogo reflexivo.